

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:43:36 ; Search time 46.76 Seconds
(without alignments)
1308.014 Million cell updates/sec

Title: US-09-757-781-21

Sequence: 1 cagcggctgctgcaggaagc.....gtctattttttttatttg 249

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/6C.COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/6D.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	36	14.5	2529	US-09-051-969A-5
2	35.2	14.1	619	US-09-328-111-579
3	31.4	12.6	579	US-09-146-950-3
4	31.4	12.6	591	US-09-146-950-19
5	31.4	12.6	1596	US-09-146-950-17
6	31.4	12.6	1724	US-08-509-024-1
7	31.4	12.6	1724	US-09-333-279-1
8	31.4	12.6	1929	US-09-146-950-1
9	31.4	12.6	4622	US-08-509-024-6
10	31.4	12.6	4622	US-09-333-279-6
11	30.4	12.2	2646	US-08-539-304A-5
12	30.4	12.2	2663	US-08-136-743B-3
13	30.4	12.2	3252	US-08-118-442-1
14	30.4	12.2	3252	US-09-677-064-1
15	29.8	12.0	1724	PCT-US96-12374-1
16	29.6	11.9	860	US-07-847-010-18
17	29.6	11.9	2658	US-08-773-608A-1
18	29.4	11.8	524	PCT-US95-08295-22
19	29.4	11.8	4875	US-08-460-739-1
20	29.2	11.7	19124	US-08-487-826B-13
21	29	11.6	279	US-08-686-878A-50
22	28.9	11.6	279	US-08-721-489-4
23	28.8	11.6	684	US-08-577-463A-1
24	28.8	11.6	2946	US-09-175-928-3
25	28.4	11.4	2622	US-08-766-014-23
26	28.4	11.4	2902	US-08-714-918-95
27	28.4	11.4	2902	US-09-265-315-95

C 28	28.4	11.4	2902	4	US-09-265-315-95	Sequence 95, Appl
C 29	28.4	11.4	2902	4	US-09-266-417-95	Sequence 95, Appl
C 30	28.4	11.4	5300	1	US-08-766-014-1	Sequence 1, Appl
C 31	28.2	11.3	618	4	US-08-961-810-17	Sequence 17, Appl
C 32	28.2	11.3	618	4	US-08-352-902D-17	Sequence 17, Appl
C 33	28.2	11.3	2075	1	US-08-238-163-3	Sequence 3, Appl
C 34	28	11.2	392	1	US-08-468-421-8	Sequence 8, Appl
C 35	28	11.2	392	1	US-08-250-975-8	Sequence 8, Appl
C 36	28	11.2	392	2	US-08-605-002A-8	Sequence 8, Appl
C 37	28	11.2	392	2	US-08-950-449A-8	Sequence 8, Appl
C 38	28	11.2	392	5	PCT-US94-10529-8	Sequence 8, Appl
C 39	28	11.2	2612	4	US-09-105-390-7	Sequence 7, Appl
C 40	28	11.2	6124	4	US-08-213-419B-3	Sequence 3, Appl
C 41	27.8	11.2	886	1	US-08-469-427A-1	Sequence 1, Appl
C 42	27.8	11.2	886	2	US-08-609-443B-1	Sequence 1, Appl
C 43	27.8	11.2	886	2	US-08-569-063C-1	Sequence 1, Appl
C 44	27.8	11.2	1002	4	US-08-960-780-43	Sequence 43, Appl
C 45	27.8	11.2	1002	4	US-09-073-898-43	Sequence 43, Appl

ALIGNMENTS

```

RESULT 1
US-09-051-969A-5
: Sequence 5, Application US/09051969A
: Patent No. 6063598
:
GENERAL INFORMATION:
: APPLICANT: ENEKEL, BARBARA
: APPLICANT: GANNON, FRANK
: APPLICANT: BERGMANN, KLAUS
: APPLICANT: NOE, WOLFGANG
: TITLE OF INVENTION: INTENSIVE HOMOLOGOUS PROMOTER OBTAINED
: TITLE OF INVENTION: FROM HAMSTERS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/051, 969A
: FILING DATE: 1998-09-30
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: FLESHNER, RAZ E.
: REGISTRATION NUMBER: 34,331
: REFERENCE/DOCKET NUMBER: 0652.1690000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2529 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2418..2465
:
US-09-051-969A-5
Query Match 14.5%; Score 36; DB 3; Length 2529;
Best Local Similarity 50.9%; Pred. No. 0.042;

```



```

: Patent No. 6287808
:
: GENERAL INFORMATION:
:
: APPLICANT: Busfield, Samantha J.
:
: TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRA-MEDIATOR-RELATED
:
: TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
:
: FILE REFERENCE: 09404/057001
:
: CURRENT APPLICATION NUMBER: US/09/146,950A
:
: CURRENT FILING DATE: 1998-09-03
:
:
: NUMBER OF SEQ ID NOS: 25
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ ID NO 17
:
: LENGTH: 1596
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: US-09-146-950-17

```

	Query Match	Best Local Similarity	12.6%;	Score 31.4;	DB 4,	Length 1596;
	Matches	74; Conservative	51.0%;	Pred. No.1;	Mismatches 71;	Indels 0;
Oy	53	gggtcgcggagatccgctcggtgctcccatcttcatttcttgtaacactcgttatgt				
Db	587	GGGTTCCTGACTACCGGAGCCTCCCTTCTGCACCCTCGGGCCGGGAAGTGCGCT				
Oy	113	aaaccatgctgtaggtgtgctaagaagtgccttgtaatcccgaatgtgaaaaaacgttgaggta				
Db	527	AAGCCGGCACAGCGGGCGCAGTGGTCCCCTCCGACGATGAGAAGTAGGCCGTGGCTGC				
Oy	173	aagctcaagcataccatgataattact				
Db	467	AgCCACACACGGCGTTCTCTGTCT				

```

RESULT      6
US-08-509-024-1/C
: Sequence 1, Application US/08509024B
: Patent No. 6291207
: GENERAL INFORMATION:
: APPLICANT: SPEAR, Patricia G.
: APPLICANT: MONTOMERY, Rebecca I.
: TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
: FILE REFERENCE: 0290-1
: CURRENT APPLICATION NUMBER: US/08/509,024B
: CURRENT FILING DATE: 1995-07-25
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 1724
: TYPE: DNA
: ORGANISM: Homo sapiens
US-08-509-024-1

```

Query Match	Similarity	12.6%	Score 31.4	DB 4	Length 1724
Best Local	Similarity	51.0%	Pred. Nm. 1.1		
Matches	74	Conservative	0	Mismatches	71
				Indels	0
				Gaps	0
Oy	53	gggtctcggagatctcggctcggctcccatcttcctatctgttcgaaacccgtattgt	112		
Db	774	gggtgtccctgactcttgggtgctctccctttgtgacacctctggccggggctggagttggc	715		
Oy	113	aaacatggtctggtgtgctaaagtgcctgtgaatcccgatgtggaanaaagctgagtgta	172		
Db	714	aagcggggcgcacggggcgagtggtcccgctcgacgacgagatgacgaagtggcctggcctgc	655		
Oy	173	aagctcagcatcacatgtatttact	197		
Db	654	agccacacacggcggttctctgtctt	630		

RESULT 7
US-09-333-279-1/c

```

: Sequence 1, Application US/09333279
: Patent No. 630336
: GENERAL INFORMATION:
: APPLICANT: SPEAR, Patricia G.
: APPLICANT: MONTGOMERY, Rebecca I.
: TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
: FILE REFERENCE: 0290-1
: CURRENT APPLICATION NUMBER: US/09/333,279
: CURRENT FILING DATE: 1999-06-15
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 1724
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-333-279-1

```

	Query Match	12.6%	Score 31.4	DB 4	Length 1724
	Best Local Similarity	51.0%	Pred. No. 1.1	71	Indels 0
	Matches 74	Conservative	0	Mismatches	Gaps
QY	53	ggagctcgagatctcgatgagctccacatctcttcattgttcgaacatccatgtaagt	112		
Db	774	ggggtgcttactatctcggtccgtcccttccgacacctctggcccgctggaagtggcgt	715		
QY	113	aaacatgagctgtygggtgactaaagtgcctctgaaatcccgatgtgaaaagctggagtgta	172		
Db	714	AAGCGGGGACGCGGGCCACAGTGTGCCCGGCTTCGAGATGCAGAAAGTGGCTTGGGCTGC	655		
QY	173	aagctcgacatacatgtaattatt	197		
Db	654	AGCCACACACGCGCTTCCTGTCTCT	630		

```

RESULT      8
US-09-146-950-1/c
: Sequence 1, Application US/09146950A
: Patent No. 6287808
: GENERAL INFORMATION:
: APPLICANT: Bushfield, Samantha J.
: TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
: TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
: FILE REFERENCE: 09404/057001
: CURRENT APPLICATION NUMBER: US/09/146,950A
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
:   LENGTH: 1929
:   TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: (297)...(875)
US-09-146-950-1

```

Query Match	12.6%	Score 31.4	DB 4	Length 1929
Best Local Similarity	51.0%	Pred. No. 1.1		
Matches	74	Conservative	0	Mismatches 71; Indels 0; Gaps 0
QY	53	gggtctcgggagctccggatcccatcttcattgtctctgaacatcctgtatgt	112	
DB	777	gggtgtcccgatcttcgggctccctcttgacacctctggcccgggctggaagtgccgt	718	
QY	113	aaacatggtctgggtgctaaagtgcctctgtaatcccgatctggaaaagctggagtgta	172	
DB	717	aagccgcggcagccggcgcgatgggtccctccctctggacgatgacgaagtggtggctgtcc	658	
QY	173	aagctcagcatatccatgtatttact	197	
DB	657	agccacacacagcgcttctgtctct	633	

```
RESULT 9
US-08-509-024-6/C
; Sequence 6, Application US/08509024B
; Patent No. 6291207
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/08/509,024B
; CURRENT FILING DATE: 1995-07-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-509-024-6
```

```
Query Match 12.6%; Score 31.4; DB 4; Length 4622;
Best Local Similarity 51.0%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 53 gggctcgggagtcctggtgctccatctccatctcattgtctgaaacatcctgtattgt 112
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 GGGTGTCTGACTCTCGGTGCTCTCTCTGACCCCTCTGCGCCGGCTGGAAGTGCGCT 485

Qy 113 aaacatgctggggtgctaaagtctgtgaatccgcatgtgaaagaagctggaggtga 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 AAGCGCGGACGCGGCGGAGTGTCGCCGCTCTGACGATGAGAAGTGCGCTGGGCTGC 425

Qy 173 aagctcagcatcattgtatttact 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 AGCCACACACGCGCTTCTCTGTCTCT 400
```

```
RESULT 10
US-09-333-279-6/C
; Sequence 6, Application US/09333279
; Patent No. 6303336
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/09/333,279
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-333-279-6
```

```
Query Match 12.6%; Score 31.4; DB 4; Length 4622;
Best Local Similarity 51.0%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 53 gggctcgggagtcctggtgctccatctccatctcattgtctgaaacatcctgtattgt 112
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 GGGTGTCTGACTCTCGGTGCTCTCTGACCCCTCTGCGCCGGCTGGAAGTGCGCT 485

Qy 113 aaacatgctggggtgctaaagtctgtgaatccgcatgtgaaagaagctggaggtga 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 AAGCGCGGACGCGGCGGAGTGTCGCCGCTCTGACGATGAGAAGTGCGCTGGGCTGC 425

Qy 173 aagctcagcatcattgtatttact 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 424 AGCCACACACGCGCTTCTCTGTCTCT 400

```
RESULT 11
US-08-539-304A-5
; Sequence 5, Application US/08539304A
; Patent No. 5792933
; GENERAL INFORMATION:
; APPLICANT: MA, DIN-POW
; TITLE OF INVENTION: FIBER-SPECIFIC PROTEIN EXPRESSION IN THE
; TITLE OF INVENTION: CORTON PLANT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/539,304A
; FILING DATE: 04-OCT-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN, OBLON F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2343-037-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(741..1093, 1220..1226)
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1094..1219
US-08-539-304A-5
```

```
Query Match 12.2%; Score 30.4; DB 1; Length 2646;
Best Local Similarity 67.2%; Pred. No. 2.8;
Matches 43; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 182 ataccatgattactacttaaaacagaaaaaaagacatgtagatagctattttt 241
    || || ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1701 ATGGGATTAATTTAAACAAAATGCTGAAAAAGAGATTATTGAGATGATTAATTTT 1760

Qy 242 ttta 245
    |||||
Db 1761 TTTA 1764
```

```
RESULT 12
US-08-136-743B-3/C
; Sequence 3, Application US/08136743B
; Patent No. 5459063
; GENERAL INFORMATION:
; APPLICANT: Barry S. Cooperman, Harvey Rubin,
; APPLICANT: Jerome Salem, and Allison L. Fisher
; TITLE OF INVENTION: "Plasmodium falciparum Ribonu-
; TITLE OF INVENTION: cleotide Reductase, DNA Sequences Therefor and Peptide inh
```


TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: The University of Pennsylvania
STREET: Suite 330
STREET: 3700 Market Street
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19104-3246
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 KB
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/136,743B
FILING DATE: 10/14/93
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 3957-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5459063e
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2663 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-136-743B-3

Query Match 12.2%; Score 30.4; DB 1; Length 2663;
Best Local Similarity 63.9%; Pred. No. 2.8;
Matches 46; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 176 ctcaagacatcgtatattcttaaaacagaaaaagacatgatagtctcat 235
DB 2301 CTTATATAAACAATGACATACATATATAATATATATATATATATATAT 2242
QY 236 tttttttttt 247
DB 2241 ATTTTTTTTTTT 2230

RESULT 13
US-09-118-442-1/C
Sequence 1, Application US/09118442B
Patent No. 6197561
GENERAL INFORMATION:
APPLICANT: Martino-Calt, Susan J.
APPLICANT: Wang, Hongyu
APPLICANT: Beach, Larry R.
APPLICANT: Wang, Xun
TITLE OF INVENTION: Genes Controlling Phylate Metabolism In
TITLE OF INVENTION: Plants and Uses Thereof
FILE REFERENCE: 0706
CURRENT APPLICATION NUMBER: US/09/118,442B
CURRENT FILING DATE: 1998-07-17
EARLIER APPLICATION NUMBER: 60/055,446
EARLIER FILING DATE: 1997-08-11
EARLIER APPLICATION NUMBER: 60/055,526
EARLIER FILING DATE: 1997-08-08
EARLIER APPLICATION NUMBER: 60/053,944
EARLIER FILING DATE: 1997-07-28
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1

LENGTH: 3252
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (258)...(2666)
NAME/KEY: misc_feature
LOCATION: (1)...(3252)
OTHER INFORMATION: n = A,T,C or G
US-09-118-442-1

Query Match 12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 92 gtctgaacatcgtatattgttaaacatgctggggtgctaaagtgcgtgaatccga 151
DB 1862 GTTCAAGCAGCTTAATAATATAGAGATCTTGAGGACACACTCTGTTAGAGAGAGT 1803
QY 152 tttggaagaaagctggagtggaagctcagcatcacatcttactttaaacagaaa 211
DB 1802 TGGTCTAATGGTGAACGAAATGGCTCATCAAGTTTGAAGCTCACTGAAGACTCTGA 1743
QY 212 aaagacatgatagtatctatttttt 243
DB 1742 TAATAGCTGCCCATTTTCAATTTCTTT 1711

RESULT 14
US-09-677-064-1/C
Sequence 1, Application US/09677064
Patent No. 6291224
GENERAL INFORMATION:
APPLICANT: Martino-Calt, Susan J.
APPLICANT: Wang, Hongyu
APPLICANT: Beach, Larry R.
TITLE OF INVENTION: Genes Controlling Phylate Metabolism In
TITLE OF INVENTION: Plants and Uses Thereof
FILE REFERENCE: 0706D
CURRENT APPLICATION NUMBER: US/09/677,064
CURRENT FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/055,446
PRIOR FILING DATE: 1997-08-11
PRIOR APPLICATION NUMBER: 60/055,526
PRIOR FILING DATE: 1997-08-08
PRIOR APPLICATION NUMBER: 60/053,944
PRIOR FILING DATE: 1997-07-28
PRIOR APPLICATION NUMBER: 09/118,442
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 3252
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (258)...(2666)
NAME/KEY: misc_feature
LOCATION: (1)...(3252)
OTHER INFORMATION: n = A,T,C or G
US-09-677-064-1

Query Match 12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 92 gtctgaacatcgtatattgttaaacatgctggggtgctaaagtgcgtgaatccga 151
DB 1862 GTTCAAGCAGCTTAATAATATAGAGATCTTGAGGACACACTCTGTTAGAGAGAGT 1807

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:33:10 ; Search time 1809.29 Seconds
(without alignments)
1857.492 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249
1 cagcggtgtgtgcaggaagc.....gtctattttttttatcgtg 249

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estnum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hnc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_hnv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	249	100.0	536	9 AI079538	AI079538 ozo4e06.x
C 2	249	100.0	536	9 AI497808	AI497808 tm89f02.x
C 3	249	100.0	561	9 AI038061	AI038061 ox21b05.x
C 4	247.4	99.4	626	10 BF679962	BF679962 602154666
C 5	242.6	97.4	618	9 BE176440	BE176440 RC3-HR058
C 6	241	96.8	540	9 AW242451	AW242451 xm899a04.x
C 7	237	95.2	486	9 AI333325	AI333325 qd05b01.x
C 8	237	95.2	466	9 AI952086	AI952086 wx46g11.x
C 9	233.2	93.7	545	9 AI129988	AI129988 qc50c03.x
C 10	233.2	93.7	687	9 BE176633	BE176633 RC3-HR058
C 11	215	86.3	463	9 AI356239	AI356239 qy65f11.x
C 12	208	83.5	462	9 AI309934	AI309934 qy01e02.x
C 13	198	79.5	446	9 AI369277	AI369277 qy91b11.x
C 14	187	75.1	436	9 AI192867	AI192867 qe68b05.x
C 15	176	70.7	427	9 AA293275	AA293275 zt38e06.s
C 16	165	66.3	179	9 AI041288	AI041288 DRFZp34M
C 17	159.2	63.9	419	9 AI739028	AI739028 w134d09.x

C 18	146.2	58.7	400	9 AI002564	AI002564 oq90g04.s
C 19	130.2	52.3	384	9 AA909465	AA909465 o14f04.s
C 20	129	51.8	378	9 AA159020	AA159020 z057c03.s
C 21	90.2	36.2	482	10 BF039164	BF039164 BP250008B
C 22	83.2	33.4	577	10 BF041786	BF041786 BP250016A
C 23	82	32.9	267	10 BE666034	BE666034 149038 MA
C 24	81.6	32.8	621	10 BF040740	BF040740 BP250010B
C 25	79	31.7	248	9 AM659322	AM659322 96381 MAR
C 26	78	31.3	599	10 BF398487	BF398487 UI-R-B52-
C 27	76.6	30.8	221	10 BM481293	BM481293 533247 MA
C 28	76.2	30.6	492	9 AA759442	AA759442 vw64f04.r
C 29	76	30.5	384	9 AI038665	AI038665 ox39f08.s
C 30	76	30.5	431	10 BG376470	BG376470 UI-R-CU0-
C 31	73	29.3	536	10 BM391560	BM391560 UI-R-DX0-
C 32	73	29.3	591	10 BM383627	BM383627 UI-R-DM1-
C 33	72	28.3	456	9 BE105642	BE105642 UI-R-BX0-
C 34	70.4	28.3	206	9 AI713888	AI713888 UI-R-AG1-
C 35	70.4	28.3	284	9 AM253128	AM253128 UI-R-BU0-
C 36	70.4	28.3	502	9 AM253386	AM253386 UI-R-BU0-
C 37	70	28.1	328	9 AI038666	AI038666 ox39f09.s
C 38	70	28.1	563	9 AI103427	AI103427 EST212716
C 39	69	27.7	444	9 AI415711	AI415711 mc74a08.x
C 40	63.6	25.5	458	9 BB830241	BB830241 BB830241
C 41	63.2	25.4	432	9 BB749365	BB749365 BB749365
C 42	62	24.9	321	9 BB837582	BB837582 BB837582
C 43	62	24.9	349	9 BF148538	BF148538 uy82e08.x
C 44	62	24.9	388	9 BB836719	BB836719 BB836719
C 45	62	24.9	411	9 AA794533	AA794533 vu68f08.r

ALIGNMENTS

RESULT 1
AI079538/c
LOCUS
DEFINITION
AI079538 536 bp mRNA linear EST 29-SEP-1998
clone IMAGE:1674370 3', mRNA sequence.

ACCESSION
AI079538
VERSION
AI079538.1 GI:3415789
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 536)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
COMMENT
This clone is available royalty-free through INLNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1447 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 445.

FEATURES

source

1. 536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1674370"
/clone_lib="Soares_fetal_liver_spleen_INFLS.S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen. Vector: p773D (Pharmacia)
with a modified polylinker; Site1: Pac I; Site2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'.
AACGGAAGATTAATTAACATCTTTTCTTTTCTTTT 3'].
double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

BASE COUNT 174 a 126 c 114 g 122 t

ORIGIN

Query Match 100.0%; Score 249; DB 9; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.1e-51;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagcggtggtgcaggaagctcactctcgctcagatagatggtgtgtggtcgcg 60
|||||
Db 497 CAGCGGTGCTGCGCAGGAAGCTCCTCGCTCAGTATTAGTGTGTGTGGGTCTCG 438
|||||
OY 61 gggatctggtgctccatctctctcattgttctgacacatctgtattgaaacatg 120
|||||
Db 437 GGGATCTCGGTGCTCCCATCTCTCTCATTTGTTCTGAACATCTGTATTGTAACATG 378
|||||
OY 121 gctgggtgtcctaaagtcctgtgaatcccgatgtggaaaaagctggagtgaaagctcag 180
|||||
Db 377 GCTGGGGTGTCTAAAGTCTGTGATCCCGATGTGAAAAAGCTGGAGGTGAACCTCAG 318
|||||
OY 181 cataccatgtattactttaaaacagaaaaagacatgatatgtatgtctattttt 240
|||||
Db 317 CATACCATGTATTACTTTAAACAGAAAAAGACATGTATGATATGCTATTATTTT 258
|||||
OY 241 ttattatcg 249
|||||
Db 257 TTTTATTGG 249

RESULT 2
AI497808 536 bp mRNA linear EST 14-APR-1999
LOCUS tm89f02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2165307 3',
DEFINITION mRNA sequence.
ACCESSION AI497808
VERSION AI497808.1 GI:4389790
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 536)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BRTP), Tumor Gene Index
Unpublished (1998)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
www-bio.llnl.gov/bdrp/image/image.html
Insert length: 779 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 468.

FEATURES
Source
Location/Qualifiers
1..536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2165307"
/clone_lib="NCI CGAP Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"

/note="Organ: brain; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGTAAGTGGAGCGGCGCCATAGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and Eco RI. The Not I and Eco RI sites of the modified p773 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 174 a 126 c 114 g 122 t

ORIGIN

Query Match 100.0%; Score 249; DB 9; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.1e-51;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagcggtggtgcaggaagctcactctcgctcagatagatggtgtgtggtcgcg 60
|||||
Db 497 CAGCGGTGCTGCGCAGGAAGCTCCTCGCTCAGTATTAGTGTGTGTGGGTCTCG 438
|||||
OY 61 gggatctggtgctccatctctcattgttctgacacatctgtattgaaacatg 120
|||||
Db 437 GGGATCTCGGTGCTCCCATCTCTCTCATTTGTTCTGAACATCTGTATTGTAACATG 378
|||||
OY 121 gctgggtgtcctaaagtcctgtgaatcccgatgtggaaaaagctggagtgaaagctcag 180
|||||
Db 377 GCTGGGGTGTCTAAAGTCTGTGATCCCGATGTGAAAAAGCTGGAGGTGAACCTCAG 318
|||||
OY 181 cataccatgtattactttaaaacagaaaaagacatgatatgtatgtctattttt 240
|||||
Db 317 CATACCATGTATTACTTTAAACAGAAAAAGACATGTATGATATGCTATTATTTT 258
|||||
OY 241 ttattatcg 249
|||||
Db 257 TTTTATTGG 249

RESULT 3
AI038061 561 bp mRNA linear EST 28-AUG-1998
LOCUS ox21b05.x1 Soares_fetal_liver_spleen_JNFs_S1 Homo sapiens cDNA
DEFINITION clone IMAGE:1656945 3', mRNA sequence.
ACCESSION AI038061
VERSION AI038061.1 GI:3277255
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 561)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1466 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 465.

FEATURES
Source
Location/Qualifiers
1..561
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1656945"
/clone_lib="Soares_fetal_liver_spleen_JNFs_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4295392"
/clone_lib="NHGMC_83"
/lab_host="DH0B (T1 phage-resistant)"
/notes="Organ: prostate; Vector: pDNR-LIB (Clontech)"

```

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-RC3-HR0585-160300-022-d046t3-2000-03-16&t4=1>)
Seq primer: puc 18 forward

High quality sequence start: 20
High quality sequence stop: 617.

FEATURES
source
Location/Qualifiers
1. 618

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0585"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pOC18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 178 a 146 c 124 g 170 t
ORIGIN

Query Match 97.4%; Score 242.6; DB 9; Length 618;
Best Local Similarity 98.4%; Pred. No. 4.2e-50;

Matches 245; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cagcgggtggtgcaggaagctcactcgcgtcagtagatgagtggtggtctcg 60
DB 303 CAGCGGTGTCGACGAGGAGCTCAGCTCGCGTCACTATGAGTGTGTGGGTCTCG 244
QY 61 gggatctgggtggtccatctctcattggttgacatccgtgttgaaacctg 120
DB 243 GGGATCTGGGTGTCGACCTCTCTCATTTCTTGAAACATCCGTATGTGAGCCATG 184
QY 121 gctgggtgctcaagtgctgtgaatccgaltgtgaaagctgaggtgaaagctcag 180
DB 183 GCTGGGTGTCGAAAGTGTGCTGATCCGATGTGAAAGCTGTGAGAGCTCG 124
QY 181 catccatgtatttactttaaaacagaaaaagacatgtagatgtagtctattttt 240
DB 123 CATACCATGTATTACTTTAAATACAGAAAAAGACATGTATGATATGTATTTT 64
QY 241 ttattatgg 249
DB 63 TTTTACTCG 55

RESULT 6
AM242451 540 bp mRNA linear EST 14-DEC-1999
LOCUS XM998404.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2692302 3',
DEFINITION mRNA sequence.
ACCESSION AM242451
VERSION AM242451.1 GI:6576205
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM at:
www-bio.llnl.gov/dbip/image/image.html
Seq primer: -400p from Glibco
High quality sequence stop: 463.

FEATURES
source
Location/Qualifiers
1. 540

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2692302"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonoids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 175 a 126 c 113 g 126 t
ORIGIN

Query Match 96.8%; Score 241; DB 9; Length 540;
Best Local Similarity 98.0%; Pred. No. 1.1e-49;

Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 cagcgggtggtgcaggaagctcactcgcgtcagtagatgagtggtggtctcg 60
DB 501 CAGCGGTGTCGACGAGGAGCTCAGCTCGCGTCACTATGAGTGTGTGGGTCTCG 442
QY 61 gggatctgggtggtccatctctcattggttgacatccgtgttgaaacctg 120
DB 441 GGGATCTGGGTGTCGACCTCTCTCATTTCTTGAAACATCCGTATGTGAGCCATG 382
QY 121 gctgggtgctcaagtgctgtgaatccgaltgtgaaagctgaggtgaaagctcag 180
DB 381 GTTGGGTGTTTAAAGTCCCTGTGAATCCGATGTGAAAGCTGTGAGAGCTCAG 322
QY 181 catccatgtatttactttaaaacagaaaaagacatgtagatgtagtctattttt 240
DB 321 CATACCATGTATTACTTTAAATACAGAAAAAGACATGTATGATATGTATTTT 262
QY 241 ttattatgg 249
DB 261 TTTTATTCG 253

RESULT 7
AI333325 486 bp mRNA linear EST 13-FEB-1999
LOCUS q905b01.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1931593
DEFINITION 3', mRNA sequence.
ACCESSION AI333325
VERSION AI333325.1 GI:4069884
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 782 Std Error: 0.00
Seq primer: -400p from Glibco
High quality sequence stop: 447.
FEATURES
source
Location/Qualifiers
1. 486
/organism="Homo sapiens"

and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo.

Query Match 93.7%; Score 233.2; DB 9; Length 545;
Best Local Similarity 98.4%; Pred. No. 9.3e-48;
Matches 246; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

BASE COUNT 172 a 128 c 116 g 129 t
ORIGIN
Query Match 93.7%; Score 233.2; DB 9; Length 545;
Best Local Similarity 98.4%; Pred. No. 9.3e-48;
Matches 246; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
Y 1 cagcgggtggtgcaggaagctcactcgcgcgctcagatagagtggtggtggtcgcg 60
Db 497 CAGCGGTGTTGGCAGGAGAGCTCAGTCCGCTCAGTATAGAGTGTGTGGGCTCG 438
Y 61 gggatcgtggtggtccatctcctt-cattgtctgaacatcctgtatgttaaacat 119
Db 437 GGGATCTGGGGCTCCATCTCTCTCCATGTTTGGACATCTGATTTAAACCAT 378
Y 120 ggcgtggtggtcgaagtcctgtgacccgagtggtgaaagctggaagtgaaagctca 179
Db 377 GCGTGGGTGCTTAAGTGTGCTGTGAATCCGATGTGGAAGAGCTGGAGGTGAAGCTCA 318
Y 180 gcatcacatgattacttacttaaaaaagaaaaagacatgatatgtctatltt 239
Db 317 GCATACCATGATTTACTTTTAAAAACAGAAAAAGACATGATGATGATGCTATTTT 258
Y 240 ttttatttgg 249
Db 257 TTTTATTGG 248

RESULT 10
LOCUS BE176633 687 bp mRNA linear EST 21-JUN-2000
DEFINITION RC3-HT0585-010400-013-b05 HT0585 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE176633
VERSION BE176633.1 GI:8639362
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 687)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jungeneel, C.V., O'Hare,
'M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and
Simpson, A.J.

Shqgun sequencing of the human transcriptome with ORF expressed
sequence tags
TITLE
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 2020263
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-RC3-HT0585-010
400-013-b05&ct3=2000-04-01&cl4=1)
Seq primer: puc 18 forward
High quality sequence start: 13
High quality sequence stop: 657.

FEATURES
SOURCE

1. 687
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone.lib="HT0585"
/dev.stage="Adult"
/note="Organ: head,neck; Vector: puc18; site.1: SmaI;
Site.2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (O.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 197 a 173 c 143 g 174 t
ORIGIN
Query Match 93.7%; Score 233.2; DB 9; Length 687;
Best Local Similarity 98.4%; Pred. No. 8.8e-48;
Matches 246; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Y 1 cagcgggtggtgcaggaagctcactcgcgcgctcagatagagtggtggtggtcgcg 60
Db 315 CAGCGGTGTTGGCAGGAGAGCTCAGTCCGCTCAGTATAGAGTGTGTGGGCTCG 256
Y 61 gggatcgtggtggtccatctcctt-cattgtctgaacatcctgtatgttaaacat 120
Db 255 GGGATCTGGGGCTCCATCTCTCTCCATGTTTGGACATCTGATTTAAACCAT 196
Y 121 gctgggtggtcgaagtcctgtgacccgagtggtgaaagctggaagtgaaagctca 180
Db 195 GCGTGGGTGCTTAAGTGTGCTGTGAATCCGATGTGGAAGAGCTGGAGGTGAAGCTCAG 136
Y 181 catcacatgattacttacttaaaaa-cagaaaaaagacatgatatgtctatltt 239
Db 135 CATACCATGATTTACTTTTAAAAACAGAAAAAGACATGATGATGATGCTATTTT 76
Y 240 ttttatttgg 249
Db 75 TTTTATTGG 66

RESULT 11
LOCUS A1356239 463 bp mRNA linear EST 16-FEB-1999
DEFINITION qy65f11.x1 NCI_CGAP_Brn25 Homo sapiens CDNA clone IMAGE:2016909 3',
mRNA sequence.
ACCESSION A1356239
VERSION A1356239.1 GI:4107860
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 463)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BRGAP), Tumor Gene Index
Unpublished (1998)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILIN at:
www.bio.liln.gov/db/brp/image/image.html
Insert length: 770 Std Error: 0.00
Seq primer: -40UP from G1bco
High quality sequence stop: 444.

FEATURES
source

1. 463
Location/Qualifiers
/organism="Homo sapiens"


```

/adb_xref="taxon:9606"
/clone_image:2016909"
/clone_id="NCI_GCAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
stranded cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAACTGAGTGGGAGCGCGGCATAGGTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

```

Query Match	86.3%;	Score 215;	DB 9;	Length 463;
Best Local Similarity	100.0%;	Pred. No. 3.1e-43;		
Matches 215;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0;
QY	35	gattacagatgctgctgctgctcgaggaatccctgctgctccatctccctcatcttt	94	
DB	463	gattatgagatgctgtctgt	404	
QY	95	ctgaacatctcttattctgtaaacatctgctgctgctgctgctgctgctgctgctgct	154	
DB	403	ctgaacatctcttattctgtaaacatctgctgctgctgctgctgctgctgctgctgct	344	
QY	155	ggaaaaaagctctgagatgtaaaagctcagcatcaccatctattactttaaaaaacgaa	214	
DB	343	ggaaaaaagctcagcatcaccatctattactttaaaaaacgaa	284	
QY	215	gacatgatatgatatgctctatttttttttattctg	249	
DB	283	gacatgatatgatatgctctatttttttttattctg	249	

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT							
AI309934/c	AI309934	AI309934	AI309934	AI309934.1	EST.	human.	Human sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 462)	NCI-CCGAP http://www.nccl.nlm.nih.gov/ncicgap .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	Unpublished (1997)	Contact: Robert Strausberg, Ph.D. Email: cgaabs-r@mail.nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.	CGNA Library Preparation: M. Bento Soares, Ph.D. CGNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert Length: 1094 Std Error: 0.00 Seq primer: -400P from Glibco	High quality sequence stop: 425.	Location/Qualifiers	1..462	/organism="Homo sapiens" /db_xref="taxon:9606"

```

/clone="IMAGE:1913978"
/clone_lib="NCI.CGAP_Kid5"
/tissue_type="2, pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AAGTGGAGAATTCGCGCGCCCAATATTTTATTTTATTTTATTTTATTTT
(Pharmacia)], digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

```

	Query Match	83.5%;	Score 208;	DB 9;	Length 462;	
	Best Local Similarity	100.0%;	Pred. No. 1.7e-41;			
	Matches	208;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
Oy	42 agtgcgtctgfgggccttcgggagatcccgcttgctcccatctctctcaattgtttcgaaca	101				
Dd	462 AGTGGTGTGTGGGCTCGGGGATCTCCGATGCCCTCCACTTTCATTGTTCGAACA	403				
Oy	102 tcccttatctgaaacatcgcctggggctgctaagtgccttgaaaccgatctgaaaaa	161				
Dd	402 TCCCTTATTGTAAACCAAGCGCTGGGGCTTAAGAAGTCCGTGAATCCCATGTGAAAAA	343				
Oy	162 gctggaggctgaagcctacagcatcacatgtattactttaaaaacagaataaagacatgt	221				
Dd	342 GCTGGAGGTGAANAAGCTCAGCATACCATTATTTACTTTAAAACAGAAAAAAGACATGT	283				
Oy	222 atggaatgctctattttttttattatgg	249				
Dd	282 ATGGAATGCTCATTTTTTTTTTATTGG	255				

```

RESULT 13
LOCUS      AI369277/c
DEFINITION 446 bp mRNA linear EST 17-MAR-1999
VERSION    4991b1.1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2019357 3'
KEYWORDS   AI369277
ACCESSION  AI369277
VERSION    AI369277.1 GI:4148030
SOURCE     EST.
ORGANISM   human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 446)
            NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTAP), "Tumor Gene Index
            Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps@email.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfield M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLT at:
            www-bio.linn.gov/bdrrp/image/image.html
            Insert Length: 723 Std Error: 0.00
            Seq primer: -40UP from Glbco.
FEATURES
            Location/Qualifiers
                1..446
                /organism="Homo sapiens"
                /db_xref="taxon:9606"

```

Query Match	79.5%	Score 198	DB 9	Length 446
Best Local Similarity	100.0%	Prod. No. 5.1e-39		
Matches 198	Conservative 0	Mismatches 0	Indels 0	Gaps 0
OY	52	tggaglcctcgggagatcctcgtcgtccatcctccctccatctgttcctgaacatccgtatcg	111	
Db	446	TGGGTCCTCGGGGATCTCGTGGCTCCCATCTTCCTTCATGTTCGAACATCCTTATGTG	387	
OY	112	taaacatcgtcgtgggtgctaaagtgcctcgtgaatcccgatgctggaataagctggaagt	171	
Db	386	TAAACCAATGCTGGGGTGTCTAAAGTGCCTGTGAATCCCGATGTGGAATAACCTGAGGTG	327	
OY	172	aaagctcagcctacatcattattcctttaaaaaacgaaaaaagcagatgatgatgct	231	
Db	326	AAAGCTCAGCATTCACATATATTATTTTAAAAACGAAAAAAGCATGTATGGATATGT	267	

```

FEATURES
source
location/Qualifiers
1..436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:174113
/clone_1ib="Soares_fetal_lung_NbHL19w"
/dev_stage="19 weeks"
/lab_host="DH10b (ampicillin resistant)"
/note="Organ: lung; Vector: pUT3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
[5'-GTGTACCAATCTGAAGTGGGAGCGGCGGCAATTTTTTTTTTTT-3']

```

Query Match	75.1%	Score 187	DB 9	Length 436
Best Local Similarity	100.0%	Pred. NO.	2.7e-36	
Matches 187; Conservative	0;	Mismatches	0;	Gaps 0

Oy	63	gatacaggctgcgtcccaatccctccatggtctgaacacctgattttaaccctgac	122
Db	436	GATCGGGTGGCTCCCATCTTCTTCATTGTTCTGAACAATCCTGATTGTAAACCATGGC	377
Oy	123	tgagggtgcataaagtgcctgtgaatcccgatgtgaaaaagcttgagggtgtaaacctcagca	182
Db	376	TGGGGTGGCTAAAGTGCCTGTGTAATCCCATGTGGAAGAAAGCTGGAGGTGAAAGCTTCACCA	317
Oy	183	taccatgtatattacttctaataaacagaaaaaaagacatgtaatgtaataatgctctatcttctt	242
Db	316	TACCATGTATTTACTTTTAAAAACAGAAAAAAGACATGTATGGAATGTCATATTTTTT	257
Oy	243	ttatttg 249	
Db	256	TTATTGG 250	

RESULT	15			
AA293275/c				
LOCUS				
DEFINITION	AA293275	427 bp	mRNA	linear
	U33806:sl Soares ovary tumor NbrHOT		Human sapiens	CDNA clone
	IMAGE:72642.3			similar to contains element LI repetitive element
				, mRNA sequence.
ACCESSION	AA293275			
VERSION	AA293275.1	GI:1941425		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Human sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 427)			
AUTHORS	Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chiapelli, B.,			
	Chisose, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins			
	, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore			
	, B., Morris, M., Parsons, J., Prange, C., Rifkin, B., Rohlfing, T.,			
	Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Travaskis, E.,			
	Underwood, K., Weidmann, P., Waterston, R., Wilson, R. and Marra, M.			
	Generation and analysis of 280,000 human expressed sequence tags			
TITLE	Genome Res. 6 (9), 807-828 (1996)			
JOURNAL				
MEDLINE	97044478			
COMMENT	Contact: Wilson RK			

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (lnl@image.llnl.gov) for further information.

Insert Length: 1397 Std Error: 0.00

Seq primer: -4lnl3 fwd. ET from Amersham

High quality sequence stop: 395.

Location/Qualifiers

1. 427

source

/organism="Homo sapiens"

/db_xref="GDB:5936269"

/db_xref="taxon:9606"

/clone="IMAGE:724642"

```

/clone_lib="Soares ovary tumor NBH07"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: ovary; Vector: pF773D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer (5'
TGTTCACATCTGAACTGGAGCGCGCGGTTTATTTTATTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pF773 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT      147 a      89 c      88 g      103 t
ORIGIN

```

```

Query Match      70.7%; Score 176; DB 9; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.5e-33;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  74  ctcccatcttccttcattgttcgaacatccgtattgttaacacatgctgggtgctaa 133
    |||||||
DB  427  ctcccatcttccttcattgttcgaacatccgtattgttaacacatgctgggtgctaa 368

QY  134  agtgcctctggaatcccgatggtggaagctggaagctgaagctcagcatatgatt 193
    |||||||
DB  367  agtgcctctggaatcccgatggtggaagctggaagctgaagctcagcatatgatt 308

QY  194  tactttaaacaagaaaaagacatgtaatgtaattttttttttttttttttttg 249
    |||||||
DB  307  tactttaaacaagaaaaagacatgtaatgtaattttttttttttttttttttg 252

```

Search completed: September 20, 2002, 05:54:01
Job time: 4851 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:48:14 ; Search time 215.36 Seconds
(without alignments)
1985.103 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249
Sequence: 1 cagcggtgtgtgcaggaagc.....gtctatctttttttatctg 249

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_032802:*

1:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	249	100.0	867	22	AAF91875	Human secreted pro
2	249	100.0	5510	22	AAH73337	Human cervical can
3	36.2	14.5	7784	24	ABL34424	Human immune syste
4	36	14.5	2529	18	AAH7468	Hamster Ubiqutin/B
5	35.4	14.2	151826	21	AAF22291	BAC containing rep
6	35.2	14.1	619	21	AAZ80495	Human colon cancer
7	34.8	14.0	1734	22	AAH17416	Human CDNA sequenc
8	34.8	13.8	2364	21	AAH70246	Plasmodium falcipa
9	34.4	13.8	15416	24	ABL34231	Human immune syste

10	34.4	13.8	15416	24	AA561453	Human gene regulat
11	34.2	13.7	7784	24	ABL34425	Human immune syste
12	34	13.7	1900	20	AAZ25197	Nucleotide sequenc
13	34	13.7	6298	22	AA545358	Chemically pretrea
14	33.8	13.6	540	22	AAH10129	Human CDNA clone (
15	33.8	13.6	5536	24	ABL32179	Human immune syste
16	33.8	13.6	7784	24	ABL34424	Human immune syste
17	33.6	13.5	13104	19	AAV52167	Streptococcus pneu
18	33.6	13.5	110000	22	AAH84800	Nucleotide sequenc
19	33.4	13.4	1267	22	AAH21012	Bovine-derived DNA
20	33.4	13.4	6306	22	AAK85746	Human immune/haema
21	33.4	13.4	8395	20	AAH3154	Enterococcus faeca
22	33	13.3	881	17	AAH40048	Human tumour necro
23	33	13.3	881	21	AAH28149	Human TR2 receptor
24	33	13.3	1704	19	AAV34509	Human TNF receptor
25	33	13.3	1704	20	AAV63763	Human tumour necro
26	32.8	13.2	265	17	AAH28133	Senescence-related
27	32.8	13.2	8032	24	AA561408	Human gene regulat
28	32.6	13.1	5914	24	ABL34169	Human immune syste
29	32.6	13.1	6398	23	ABL13526	Drosophila melanog
30	32.4	13.0	5062	22	AA546696	Tumour suppressor
31	32.2	12.9	500	21	AAH5018	Cat flea head and
32	32.2	12.9	1804	23	AAH3281	DNA encoding novel
33	32.2	12.9	3069	22	AAH54988	S. epidermidis gen
34	32.2	12.9	3206	22	AAH55018	S. epidermidis gen
35	32.2	12.9	4249	22	AAH54769	S. epidermidis gen
36	32.2	12.9	5536	24	ABL32178	Human immune syste
37	32.2	12.9	6351	24	ABL33745	Human immune syste
38	32.2	12.9	6351	24	ABL34587	Human immune syste
39	32	12.9	5536	24	ABL32179	Human metastasis a
40	32	12.9	17341	21	AAH14872	Human immune syste
41	32	12.9	32150	22	AA532454	Genomic DNA sequen
42	31.8	12.8	407	22	AAH20983	Human genomic DNA
43	31.8	12.8	407	22	AAH69467	Human nervous syst
44	31.8	12.8	1383	24	AAH59277	Human digestive sy
45	31.8	12.8	5314	24	ABL32160	Long terminal repe
						Human immune syste

ALIGNMENTS

RESULT	1
AAF91875	AAF91875 standard; CDNA; 867 BP.
ID	AAF91875;
XX	XX
AC	AAF91875;
XX	XX
DT	22-MAY-2001 (first entry)
XX	XX
DE	Human secreted protein-encoding gene 18 CDNA clone HEPF57, SEQ ID NO:28.
XX	XX
KW	Human; secreted protein; proliferative disorder; cancer; tumour;
KW	focal abnormality; developmental abnormality; haematopoietic disorder;
KW	immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;
KW	inflammation; allergy; neurological disorder; Alzheimer's disease;
KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW	cardiovascular disorder; angiotensin disorder; kidney disorder;
KW	gastrointestinal disorder; pregnancy-related disorder;
KW	endocrine disorder; infection; wound healing; vulnerrary;
KW	cell culture; chemotaxis; food additive;
KW	binding partner identification; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200118022-A1.
XX	XX
PD	15-MAR-2001.
XX	XX
PF	31-AUG-2000; 2000OMO-US24008.
XX	XX
PR	03-SEP-1999; 99US-0152315.
PR	03-SEP-1999; 99US-0152317.

XX (HUMA-) HUMAN GENOME SCI INC.
PA

XX N1 J. Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX

XX WPI: 2001-203081/20.
DR P-PSDB; AAB87359.
XX

PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX

XX Claim 1; Page 498; 607pp; English.
XX

CC AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
CC
XX

SQ Sequence 867 BP; 233 A; 180 C; 197 G; 256 T; 1 other;

Query Match 100.0%; Score 249; DB 22; Length 867;
Best Local Similarity 100.0%; Pred. No. 2.9e-65;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtggtctcg 60
DB 348 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtggtctcg 407
QY 61 gggatctcgggtggtcgcacatcttcctcattgtctgaaacatccgtatgtaaacatg 120
DB 408 gggatctcgggtggtcgcacatcttcctcattgtctgaaacatccgtatgtaaacatg 467
QY 121 gctgggggtgctaaagtgctctgtaatcccgatggtggaagtcggaggtggaagtcag 180
DB 468 gctgggggtgctaaagtgctctgtaatcccgatggtggaagtcggaggtggaagtcag 527
QY 181 cataccatgattacttacttaaaacagaaaaagacatgatatgatatcttttt 240
DB 528 cataccatgattacttacttataaaacagaaaaagacatgatatgatatcttttt 587
QY 241 ttattatgg 249
DB 588 ttattatgg 596

RESULT 2

AAH73337
ID AAH73337 standard; cDNA; 5510 BP.
XX

XX AAH73337;
AC
XX

XX 19-SEP-2001 (first entry)
DE
XX

XX Human cervical cancer marker nucleic acid 4611.
XX
XX

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
XX

OS Homo sapiens.
XX
XX

XX WO200142467-A2.
XX
XX

XX 14-JUN-2001.
XX
XX

XX 08-DEC-2000; 2000WO-US33312.
XX
XX

XX 08-DEC-1999; 99US-0169681.
XX
XX

XX 21-DEC-1999; 99US-0171350.
XX
XX

XX 14-MAR-2000; 2000US-0189315.
XX
XX

XX 12-MAY-2000; 2000US-0203791.
XX
XX

XX 09-JUN-2000; 2000US-0210600.
XX
XX

XX 21-JUL-2000; 2000US-0220114.
XX
XX

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX

XX Schlegel R, Deeds J, Berger A, Zhao X;
XX
XX

XX WPI: 2001-375006/39.
XX
XX

XX New isolated nucleic acid for diagnosing and treating cervical cancer
XX and for assessing and detecting compounds for treating the cancer -
XX
XX

XX Claim 1; Page 1029-1031; 1051pp; English.
XX
XX

XX The invention relates to novel genes (AAH68727-AAH73383) associated with
XX cervical cancer with cytostatic activity. The nucleic acids and encoded
XX polypeptides are useful to assess if a patient is afflicted with
XX cervical cancer or has a pre-malignant condition; to monitor the
XX progression of cervical cancer or a premalignant condition in a patient;
XX and to select and/or assess the efficacy of a compound or therapy for
XX inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy.
XX

SQ Sequence 5510 BP; 1526 A; 1242 C; 1418 G; 1305 T; 19 other;

Query Match 100.0%; Score 249; DB 22; Length 5510;
Best Local Similarity 100.0%; Pred. No. 5.7e-65;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtggtctcg 60
DB 4914 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtggtctcg 4973
QY 61 gggatctcgggtggtcgcacatcttcctcattgtctgaaacatccgtatgtaaacatg 120
DB 4974 gggatctcgggtggtcgcacatcttcctcattgtctgaaacatccgtatgtaaacatg 5033
QY 121 gctgggggtgctaaagtgctctgtaatcccgatggtggaagtcggaggtggaagtcag 180
DB 5034 gctgggggtgctaaagtgctctgtaatcccgatggtggaagtcggaggtggaagtcag 5093
QY 181 cataccatgattacttacttaaaacagaaaaagacatgatatgatatcttttt 240
DB 5094 cataccatgattacttacttataaaacagaaaaagacatgatatgatatcttttt 5153
QY 241 ttattatgg 249
DB 5153 ttattatgg 5153

Db	412	CAGTAAATCTGGCTTTTACATGCATCTCTACGAAAGTACGTAAGCTGTGCAGTTACTTCTC	413
Oy	199	taaaaaacagaaaaaagacatgcatgatatgctacttttttt	243
Db	412	TTAAACACACAGAAAAACGTATATTGGCAGATCTATGTTT	368
RESULT	7		
AAH17416			
AAH17416		standard; cDNA; 1734 BP.	
XX			
AC	AAH17416:		
XX			
DT	26-JUN-2001	(first entry)	
XX			
DE	Human cDNA sequence SEQ ID NO:16860.		
XX			
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss		
XX			
OS	Homo sapiens.		
XX			
PN	EP1074617-A2.		
XX			
PD	07-FEB-2001.		
XX			
PF	28-JUL-2000; 2000EP-0116126.		
XX			
PR	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,		
PI	Isht S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
DR	WPI; 2001-318749/34.		
XX			
PT	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	full-length cDNAs defined in the specification, and for the detection		
PT	and/or diagnosis of the abnormality of the proteins encoded by the		
PT	full-length cDNAs -		
PS	Claim 8; SEQ ID 16860; 2537bp + CD ROM; English.		
XX			
CC	The present invention describes primer sets for synthesizing 5602		
CC	full-length cDNAs defined in the specification. Where a primer set		
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
CC	to the complementary full-length cDNA; and (b) a primer complementary		
CC	to the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in		
CC	the specification. The primer sets can be used in antisense therapy and		
CC	in gene therapy. The primers are useful for synthesizing polynucleotides,		
CC	particularly full-length cDNAs. The primers are also useful for the		
CC	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	the full-length cDNAs. The primers allow obtaining of the full-length		
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and		
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to		
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632		
CC	represent oligonucleotides, all of which are used in the exemplification		
XX	of the present invention.		
XX			
XX	Sequence 1734 BP; 371 A; 484 C; 428 G; 451 T; 0 other;		

	Query Match	14.0%	Score 34.8;	DB 22;	Length 1734;
	Best Local Similarity	53.7%;	Fred. No. 1.5;		
	Matches 72;	Conservative 0;	Mismatches 62;	Indels 0;	Gaps 0;
Oy	115 acccagtcgctgggtgcctaagtcccttgatccccgaatgtagaaaaagcttggaa 174 				
Db	1293 accctcgacttggaattcatgatctggaataactacagtaaaccaaatgycagaag 1352				
Oy	175 gctcagcatlacatgttatctttaaaaacagaaaaagacatgtatgatalgtcta 234 				
Db	1353 ccaggtctcaggttatatgtacctgaagctctgtattaagaattctctcttttct 1412				
Oy	235 tttttttttatgt 248 				
Db	1413 ttttttttttgt 1426				
RESULT 8					
ID	AAA70246/c				
XX	AAA70246 standard; DNA; 2364 BP.				
AC	AAA70246;				
XX					
DT	07-NOV-2000 (first entry)				
XX					
DE	Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:379.				
KW	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacidie; infection; insecticide; ds.				
XX					
OS	Plasmodium falciparum.				
PX	WO200025728-A2.				
PD					
XX	11-MAY-2000.				
FE	05-NOV-1999; 99MO-US26796.				
PR					
XX	05-NOV-1998; 98US-0107131.				
PA	(HOEF/) HOFFMAN S. (CARU/) CARUCCI D. (GARD/) GARDNER M. (VENT/) VENTER J C.				
Hoffman S,	Carucci D,	Gardner M,	Venter JC:		
WP1:	2000-365347/31.				
Proteins encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum, useful as antimalarial vaccines and in the diagnosis of P.falciparum infection -					
Disclosure: Page 558-559; 577pp; English.					
The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum. Also described are: (I) nucleotide sequences (II) encoding (I); and (2) vaccines against P. falciparum infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against P. falciparum infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with P. falciparum. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent P. falciparum infection, or they can be used to identify drug resistance in P. falciparum. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new					

CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 2364 BP; 1222 A; 247 C; 213 G; 682 T; 0 other;

Query Match	14.0%;	Score 34.8;	DB 21;	Length 2364;
Best Local Similarity	65.4%;	Pred. No. 1.7;		
Matches 51;	Conservative 0;	Mismatches 27;	Indels 0;	Gaps 0;

Qy 170 tgaagccagcataccatgtaatttaactttaaaaaacagaaaaaacatgtaagatat 229
 |||| | |||| | |||| |||| | |||| | |||| | ||||
 Db 1854 TGAATTATCTGCATGAGATGACCTTAATTATTATCATCAACAATAACATTTTTTGATAT 1795

```

Qy      230  gtctatttttttatt 247
          | | | | | | | |
Db      1794 ATGTGAATTATTTTATT 1777

```

RESULT	9
ABL34231	
ID	ABL34231 standard; DNA; 15416 BP.

Human Immune system associated gene SEQ ID NO: 2204.

KW Human; immune system disease;cyclosine methylation; antisthmatic;
KW antiarteriosclerotic; antianemic; cytosolic; nootropic;
KW neuroprotective; anti-HIV; antiviral; ophthalmological;
KW antihypertensive; antithyroidic; antidiabetic; antipsoriatic;
KW antifungal; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; Rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

OS	Homo sapiens.
XX	
PN	WO200200928-A2.

PD 03-JAN-2002.

02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.

PA (EPiG-) EPiGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -

PS Claim 1; SEQ ID NO 2204; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and, macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

SQ Sequence 15416 BP; 4742 A; 239 C; 3068 G; 7367 T; 0 other;

Query Match	13.8%;	Score 34.4;	DB 24;	Length 15416;
Best Local Similarity	57.4%;	Pred. No. 4.5;		
Matches 62;	Conservative 0;	Mismatches 46;	Indels 0;	Gaps 0;

Qy 140 tgggaatcccgatgttgaaanaagctcgaggtgaaagctcagcatcatgtattacttt 199
 | | | | | | | | | | | | | | | | | |
 Db 1966 tttaatttttgatgatgcaattataataaatgtatagatagaatgtcttaagtt 2025

QY	200	aaaaacagaaaaaaagacatgtagtagatactgtctatttttttttatt	247
Db	2026	tataaagaatacaagtatatgatgaggtctcttttttttttttttttt	2073

RESULT 10
AAS61453
ID AAS61453 standard; DNA; 15416 BP.

DT	29-JAN-2002	(first entry)
XX		
DE	Human gene regulation-associated gene oligonucleotide #408	

KM Human; Gene regulation-associated gene; severe combined immunodeficiency
KM cardiac damage; inflammatory response; Hemophilia; Werner syndrome
KM asthma; HFR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KM renal disease; Preeclampsia; cardiac allograft vascular disease;
KM colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KM immunostimulant; cardiant; antineoplastic; antihistaminic;
KM neurotrophic; gynecological; anti-tumour; immunosuppressive; cytostatic.
KM

OS	Homo sapiens.
XX	
PN	WO200177375-A2.

PD 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP039668.

PR 06-APR-2000; 2000DE-1019058.

PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-017470/02

PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -

PS Disclosure; SEQ ID No 418; 26pp; English

The invention relates to 224 nucleic acid sequences comprising at least 18 bases of a chemically pretreated gene associated with gene regulation selected from 43 known genes (or complementary sequences). The chemical pretreatment converts cytosine bases into thymine at the 5-position to uracil or another base with hybridisation behaviour dissimilar to cytosine, to enable analysis of cytosine methylations. The DNA sequences, oligomers (or sets/arrays) and method are useful in the diagnosis of diseases (or predisposition to diseases) associated with gene regulation and in therapy of such diseases, by enabling analysis of the cytosine methylation patterns of such genes, kits are provided. They are especially useful in diagnosis and therapy of e.g. severe combined immunodeficiency disease, cardiac disorders, haemophilia, solid tumours and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease, pre-eclampsia, graft versus-host disease. The present sequence is a

Db 1891 TGTAATAATGATGATATTTCAAAAATCTATTATTAACAGCTTCAATAAATTACAG 1832
 Oy 170 tgaagcttcgacacacatgattactttaaaacagaaaaagacatgtagat 229
 Db 1831 ATAAAGTACATTTTGTGCTGATTAATAAACAACGATTAAGAAAACAATATTTT 1772
 Oy 230 gtcatcttttttatt 247
 Db 1771 CATAAATGCTGTTTATTT 1754

RESULT 13

AAS45358

ID AAS45358 standard; DNA; 6298 BP.

XX AAS45358;

XX 18-DEC-2001 (first entry)

XX Chemically pretreated genomic DNA associated with cell cycle #32.

DE Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KM PCR primer.

XX Homo sapiens.

XX WO200168911-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02945.

XX 15-MAR-2000; 2000DE-1013847.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIS-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2001-602751/68.

XX Designing primers and probes for analysing diseases associated with

XX cytosine methylation state e.g. arthritis, cancer, aging,

XX arteriosclerosis comprising fragments of chemically modified genes

XX associated with cell cycle -

XX Claim 1; SEQ ID No 63; 28pp; English.

PS Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all Cpg dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences and/or the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers.

XX Sequence 6298 BP; 1583 A; 182 C; 1402 G; 3131 T; 0 other;
 XX

Query Match 13.7%; Score 34; DB 22; Length 6298;
 Best Local Similarity 56.1%; Pred. No. 4.3;
 Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 Oy 134 agtgccttgatccctgttgaaagcctgaggggtgaagctcagcaccatgatt 193
 Db 3416 agtgccttgatccctgttgaaagcctgaggggtgaagctcagcaccatgatt 3475
 Oy 194 tacttaaaaaacagaaaaagacatgtagatgtagatgtagatgtagatgtagat 247
 Db 3476 tttttataataaaaaaagaattatgtagatgtagatgtagatgtagatgtagat 3529

RESULT 14

AAH10129/C

ID AAH10129 standard; cDNA; 540 BP.

XX AAH10129;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (3'-primer) SEQ ID NO:6964.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 3; SEQ ID 6964; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:01:20 : Search time 2113.79 seconds
(without alignments)
2465.102 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249
Sequence: 1 cagcgcgtgctgcagcgaagc.....gtctatctttttttattg 249

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sy:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hcgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Match Length DB ID Description

1	249	100.0	2696	9	AK000761	AK000761 Homo sapi
2	249	100.0	5958	9	AF196185	AF196185 Homo sapi
3	249	100.0	22279	2	U82210	U82210 Homo sapien
4	249	100.0	106198	9	AL160409	AL160409 Human DNA
5	249	100.0	138688	2	AC025823	AC025823 Homo sapi
6	249	100.0	196374	2	AL591464	AL591464 Homo sapi
7	176	70.7	427	6	AX332939	AX332939 Sequence
8	70	28.1	5500	10	AB005549	AB005549 Rattus no
9	50.6	20.3	168213	2	AC103175	AC103175 Rattus no
10	40.8	16.4	168280	2	AC103460	AC103460 Rattus no
11	39	15.7	2305	3	S96842	S96842 SP96-spore-
12	38.8	15.6	176741	9	AC016732	AC016732 Homo sapi
13	38.6	15.5	203194	2	AC023228	AC023228 Homo sapi
14	38.4	15.4	119579	2	AC008425	AC008425 Homo sapi
15	38.2	15.3	92795	2	AL357035	AL357035 Homo sapi
16	38.2	15.3	189655	2	AC103542	AC103542 Rattus no
17	38	15.3	197239	2	AC094815	AC094815 Rattus no
18	37.6	15.1	162260	2	AC021236	AC021236 Homo sapi
19	37.4	15.0	55774	2	AC091303	AC091303 Homo sapi
20	37.2	14.9	76113	2	AC023453	AC023453 Homo sapi
21	37	14.9	125301	9	HS388N15	Z99571 Human DNA s
22	36.8	14.8	12029	3	AE001412	AE001412 Plasmodiu
23	36.8	14.8	185182	9	AC093423	AC093423 Homo sapi
24	36.6	14.7	39608	9	AC009004	AC009004 Homo sapi
25	36.6	14.7	74470	9	AC079404	AC079404 Homo sapi
26	36.6	14.7	79085	9	AC006221	AC006221 Homo sapi
27	36.6	14.7	164309	30	AC067834	AC067834 Homo sapi
28	36.6	14.7	166500	9	AC007785	AC007785 Homo sapi
29	36.6	14.7	169572	2	AC108219	AC108219 Homo sapi
30	36.6	14.7	191830	2	AC025660	AC025660 Homo sapi
31	36.6	14.7	193351	9	AC025656	AC025656 Homo sapi
32	36.6	14.7	205908	8	AL669903	AL669903 Mus muscu
33	36.4	14.6	34722	2	SPBC16C6	AL021767 S. pombe c
34	36.4	14.6	36629	2	AC107279	AC107279 Rattus no
35	36.4	14.6	186135	2	AL669927	AL669927 Mus muscu
36	36.2	14.5	110000	2	AX347326	AX347326 Sequence
37	36.2	14.5	110000	2	AP002753.2	Continuation (3 of
38	36.2	14.5	110000	10	AE008684.1	Continuation (2 of
39	36.2	14.5	131972	9	AC008794	AC008794 Homo sapi
40	36.2	14.5	132090	10	AC005402	AC005402 Mus muscu
41	36.2	14.5	151558	9	AC021766	AC021766 Homo sapi
42	36.2	14.5	155633	2	AC078795	AC078795 Homo sapi
43	36.2	14.5	156060	2	AC004153	AC004153 Plasmodiu
44	36.2	14.5	169346	2	AC004157	AC004157 Plasmodiu
45	36.2	14.5	169794	2	AC004688	AC004688 Plasmodiu

ALIGNMENTS

RESULT	1	2696 bp	mRNA	Linear	PRI 22-PEB-2000
LOCUS	AK000761				
DEFINITION	Homo sapiens cDNA FLJ20754 fis, clone HEP02246.				
ACCESSION	AK000761				
VERSION	AK000761.1 GI:7021050				
KEYWORDS	Oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens hepatoma cell_line:Hep2 cDNA to mRNA, clone_lib:HEP clone:HEP02246.				
ORGANISM	Homo sapiens				
REFERENCE					
AUTHORS	1 (sites)				
	Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,				
	Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,				
	Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,				
	Nakamura,Y., Isogal,T., and Sugano,S.				
	NEDO human cDNA sequencing project				
TITLE	Unpublished (2000)				
JOURNAL	2 (bases 1 to 2696)				
REFERENCE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogal,T.,				
AUTHORS	Shibahara,T., Tanaka,T. and Nakamura,Y.				
TITLE	Direct Submission				


```

Db 5582 GCTGGGGTCTAAGTCCGTGATCCGATGTGGAAGAGCTGGAGGTGAAGCTCAG 5641
QY 181 catcacatgtattactttaaaacagaaacacatgtatgatatgtctattttt 240
Db 5642 CATACCATGTATTACTTTAAACAGAAAAAAGACATGTATGATGTCTATTTT 5701
QY 241 ttattttg 249
Db 5702 TTTTATTGG 5710

RESULT 3
LOCUS U82210 22279 bp DNA linear HTG 30-AUG-2001
DEFINITION Homo sapiens chromosome 10 clone CRI-JC2075 map 10p11.2, ***
ACCESSION U82210
VERSION U82210.1 GI:3983510
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 22279)
AUTHORS Smith,D.R.
TITLE Sequencing of Human Chromosome 10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 22279)
AUTHORS Du,L. and Smith,D.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1996) Bioinformatics Division, Genome
Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
On Dec 9, 1998 this sequence version replaced gi:1773048.
COMMENT Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
Technology. Data may contain low quality sequence and BAC/Cosmid
vector sequences.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 22279: contig of 22279 bp in length.
FEATURES
source 1..22279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
/clone="CRI-JC2075"

BASE COUNT 5671 a 5238 c 5397 g 5959 t 14 others
ORIGIN

Query Match 100.0%; Score 249; DB 2; Length 22279;
Best Local Similarity 100.0%; Pred. No. 1.8e-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcggatggtgcaggaagtcacatctcgcgtcagatattagagtgtgtgtgtcgcg 60
Db 10975 CAGCGGTGTGTGCGAGACACTCCTCGCGTCAATTAGACTGTGTGTGGTCTCG 11034
QY 61 gggatctcgtgtgtcctccatctctctcattgtctcgaacatctgtattgttaaccatg 120
Db 11035 GGGATCTCGTGTGCTCCATCTTCCTTCATTTGTCGAAACATCCTGATATGTAACCATG 11094
QY 121 gctgggggtctaaagtgtcgtatgcacccgcatgtgaaaaagcttgagtgaaagctcag 180
Db 11095 GCTGGGGTCTAAAGTGTCTGTGATCCGATGTGGAAGAGCTGAGAGTCTCAG 11154

```

```

QY 181 catcacatgtattactttaaaacagaaacacatgtatgatatgtctattttt 240
Db 11155 CATACCATGTATTACTTTAAACAGAAAAAAGACATGTATGATGTCTATTTT 11214
QY 241 ttattttg 249
Db 11215 TTTTATTGG 11223

RESULT 4
LOCUS AL160409/c 106198 bp DNA linear PRI 18-JUL-2001
DEFINITION Human DNA sequence from clone RP11-406D17 on chromosome 10,
complete sequence.
ACCESSION AL160409
VERSION AL160409.12 GI:14970794
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 106198)
AUTHORS Sycamore,N.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jul 19, 2001 this sequence version replaced gi:14625535.
COMMENT During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-406D17 is from the library RP11-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-406D17 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP13-348N17 is at 104199 in this
sequence. The true right end of clone RP11-490024 is at 2000 in
this sequence.
FEATURES
source 1..106198
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-406D17"
/clone_1lb="RP11-11.2"
27..160
/note="MIR repeat: matches 3..139 of consensus"
repeat_region 989..1039
/note="MIR repeat: matches 93..144 of consensus"
repeat_region 1232..1275
/note="22 copies 2 mer ac 77% conserved"

```

repeat_region	1344. .1651	/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region	4229. .4606	/note="AluSx repeat: matches 1. .403 of consensus"
repeat_region	4648. .4815	/note="MLTJ repeat: matches 1. .403 of consensus"
repeat_region	5215. .5516	/note="FRAM repeat: matches 4. .171 of consensus"
repeat_region	6161. .6216	/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region	6217. .6527	/note="MIR repeat: matches 47. .104 of consensus"
repeat_region	6528. .6637	/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region	6646. .6950	/note="MIR repeat: matches 104. .205 of consensus"
repeat_region	8001. .8354	/note="AluDb repeat: matches 1. .304 of consensus"
repeat_region	8443. .8602	/note="AluDb repeat: matches 1. .289 of consensus"
repeat_region	10005. .10072	/note="AluIo/FRAM repeat: matches 152. .308 of consensus"
repeat_region	10342. .10849	/note="L2 repeat: matches 2673. .2741 of consensus"
repeat_region	10851. .10975	/note="MER57B repeat: matches 43. .386 of consensus"
repeat_region	11451. .11758	/note="FLAMC C repeat: matches 15. .133 of consensus"
repeat_region	12392. .14459	/note="AluIo repeat: matches 1. .312 of consensus"
repeat_region	14460. .14550	/note="pR5 repeat: matches 504. .1874 of consensus"
repeat_region	14551. .14771	/note="MER61E repeat: matches 377. .466 of consensus"
repeat_region	14772. .14923	/note="pR5 repeat: matches 1874. .8522 of consensus"
repeat_region	14930. .15218	/note="AluSx repeat: matches 3. .154 of consensus"
repeat_region	15223. .15519	/note="AluDb repeat: matches 1. .293 of consensus"
repeat_region	15661. .15818	/note="AluSg repeat: matches 1. .299 of consensus"
repeat_region	15868. .15919	/note="MER5A repeat: matches 34. .189 of consensus"
repeat_region	16133. .16443	/note="26 copies 2 mer aa 73% conserved"
repeat_region	16599. .16707	/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region	16712. .16787	/note="LMC4 repeat: matches 7728. .7834 of consensus"
repeat_region	16825. .17092	/note="38 copies 2 mer ta 76% conserved"
repeat_region	17094. .17254	/note="AluIo repeat: matches 1. .292 of consensus"
repeat_region	17579. .17889	/note="AluIo/FRAM repeat: matches 134. .295 of consensus"
repeat_region	18807. .19086	/note="AluSg repeat: matches 1. .306 of consensus"
repeat_region	19087. .19391	/note="MLTJ repeat: matches 3. .238 of consensus"
repeat_region	19392. .19542	/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region	19608. .19904	/note="MLTJ repeat: matches 238. .501 of consensus"
repeat_region	20333. .20642	/note="AluIo repeat: matches 1. .295 of consensus"
repeat_region	20670. .20956	/note="AluSx repeat: matches 1. .310 of consensus"
repeat_region	21297. .21601	/note="AluDb repeat: matches 1. .286 of consensus"
repeat_region	21857. .22175	/note="L2 repeat: matches 2414. .2696 of consensus"
repeat_region	22502. .22739	/note="AluIo repeat: matches 1. .311 of consensus"

repeat_region	/note="AluX repeat: matches 1. .239 of consensus" 23135. .23915
repeat_region	/note="L1PA10 repeat: matches 5369. .6158 of consensus" 27500. .27629
repeat_region	/note="L1MA3 repeat: matches 6171. .6304 of consensus" 2661. .28984
repeat_region	/note="AluSg repeat: matches 1. .312 of consensus" 30001. .30132
repeat_region	/note="66 copies 2 mer aa 62% conserved" 30233. .30328
repeat_region	/note="48 copies 2 mer ta 65% conserved" 30365. .30644
repeat_region	/note="L1MC4 repeat: matches 6633. .6940 of consensus" 30645. .31062
repeat_region	/note="MSTA repeat: matches 1. .424 of consensus" 31063. .31122
repeat_region	/note="L1MC4 repeat: matches 6940. .6995 of consensus" 31133. .31547
repeat_region	/note="L1MC4 repeat: matches 7167. .7597 of consensus" 31648. .32167
repeat_region	/note="L1PA6 repeat: matches 5612. .6131 of consensus" 32275. .32552
repeat_region	/note="AluJo repeat: matches 5. .282 of consensus" 32592. .32686
repeat_region	/note="L1MC4 repeat: matches 7791. .7972 of consensus" 33154. .33272
repeat_region	/note="MIR repeat: matches 80. .185 of consensus" 33283. .33415
repeat_region	/note="AluJo repeat: matches 14. .153 of consensus" 33416. .34701
repeat_region	/note="pR5 repeat: matches 172. .2212 of consensus" 34713. .34840
repeat_region	/note="AluJo/FRAM repeat: matches 178. .302 of consensus" 35365. .35654
repeat_region	/note="AluJo repeat: matches 1. .288 of consensus" 36067. .36155
repeat_region	/note="I2 repeat: matches 2622. .2710 of consensus" 36582. .36804
repeat_region	/note="MER30 repeat: matches 1. .230 of consensus" 37547. .37852
repeat_region	/note="AluX repeat: matches 1. .311 of consensus" 37894. .38050
repeat_region	/note="MER5A repeat: matches 9. .188 of consensus" 38447. .38744
repeat_region	/note="AluX repeat: matches 5. .302 of consensus" 39063. .39193
repeat_region	/note="FRAM_C repeat: matches 1. .131 of consensus" 39510. .39673
repeat_region	/note="MER5A repeat: matches 4. .189 of consensus" 40932. .41102
repeat_region	/note="L1PA7 repeat: matches 1497. .1667 of consensus" 41103. .41726
repeat_region	/note="L1PA7 repeat: matches 5515. .6138 of consensus" 43386. .43747
repeat_region	/note="THE1B repeat: matches 1. .364 of consensus" 44459. .44771
repeat_region	/note="AluSg repeat: matches 3. .306 of consensus" 44991. .45302
repeat_region	/note="AluJo repeat: matches 1. .304 of consensus" 45536. .51640
repeat_region	/note="L1PA7 repeat: matches 11. .6145 of consensus" 51663. .51964
repeat_region	/note="AluB repeat: matches 1. .308 of consensus" 51966. .52183
repeat_region	/note="109 copies 2 mer aa 69% conserved"

Query Match	100.0%	Score 249	DB 9	Length 106198
Best Local Similarity	100.0%	Pred. No. 2e-60		
Matches 249	Conservative	0	Mismatches	0
			Indels	Gaps
QY	1	cagcgctgtgtgcaggaacctcactctcgcgcgcagtatagatgtgtgtgtgctgcg	60	
DB	88101	CAGCGGTGTGTGCAGGAACCTCACACTCTCGCGCAGTATTAGATGTGTGTGGGCTCG	88042	


```

OY      61  ggagctcgtggtcctccatcttccttcattgtctcgaacatctgtattgttaaccatg 120
          |||||||
Db      88041  GGGATCTCGGTGGCTCCATCTTCATTGTTCTGAACATCTGTATGTAAACCATG 87982
OY      121  gctgggggtcctaagtgctgtaacccgagctggaagagctggaagcctcag 180
          |||||||
Db      87981  GCTGGGCTCTTAAGTCTGTGATCCGATGTGAAAAAGCTGGAGGGAAGCTCAG 87922
OY      181  cataccatgtattactttaaaaaacagaaaacagacatgatagtatgtctattttc 240
          |||||||
Db      87921  CATACCATGTATTACTTAAAAACAGAAAAAAGACATGTATGATATCTATTTT 87862
OY      241  ttattattg 249
          |||||||
Db      87861  TTTTATTGG 87853

RESULT  5
AC025823 LOCUS      138688 bp      DNA      linear      HTG 24-JAN-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-60H16, WORKING DRAFT
ACCESSION  AC025823
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 138688)
            Smith,D.R.
            Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
            Unpublished
            2 (bases 1 to 138688)
            Smith,D.R.
            Direct Submission
            Submitted (16-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
            On Apr 27, 2001 this sequence version replaced gi:987686.
            -----
            Genome Center
            Center: Genome Therapeutics Corporation
            Web site: http://www.genomecorp.com/
            Center code: GTC
            Contact: gtc-seqcenter@genomecorp.com
            -----
            Project Information
            Center project name: hg317
            -----
            Summary Statistics
            Sequencing vector: N/A
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 990315
            Consensus quality: 133029 bases at least Q40
            Consensus quality: 135140 bases at least Q30
            Consensus quality: 136233 bases at least Q20
            Insert size: 138187; sum-of-ctrls
            Quality coverage: 6.3x in Q20 bases; sum-of-ctrls
            -----
            NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1
            1095: contig of 1095 bp in length
            *
            1196: gap of unknown length
            *
            1196: contig of 1008 bp in length
            *
            2303: gap of unknown length
            *
            2304: contig of 7159 bp in length
            *
            9462: gap of unknown length
            *
            9562: contig of 16245 bp in length
            *
            9563

```

```

FEATURES
source
    1. 138688
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="10"
        /clone="RP11-60H16"
        /clone_id="RP11-11"
        1..1095
        /note="assembly_name:Contig14"
        1196..2203
        /note="assembly_name:Contig15"
        2304..9462
        /note="assembly_name:Contig55"
        /clone_end:"7"
        9563..25807
        /note="assembly_name:Contig56"
        /clone_end:"SP6"
        25808..39871
        /note="assembly_name:Contig57"
        39872..68325
        /note="assembly_name:Contig58"
        68426..138688
        /note="assembly_name:Contig59"
BASE COUNT  41112 a 30000 c 29755 g 37221 t      600 others
ORIGIN
Query Match      100.0%; Score 249; DB 2; Length 138688;
Best Local Similarity 100.0%; Pred. No. 2; ie-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1  cagcggtggtgagcagaagctcctcgcgcagatagagtggtggtcgcg 60
          |||||||
Db      134742  CAGCGGTGTTGGGAGAGGCTCCTCGCTCAGTATTAGTGTGTGGGCTCG 134683
OY      61  ggagctcgtggtcctccatcttccttcattgtctcgaacatctgtattgttaaccatg 120
          |||||||
Db      134682  GGGATCTCGGTGGCTCCATCTTCATTGTTCTGAACATCTGTATGTAAACCATG 134623
OY      121  gctgggggtcctaagtgctgtaacccgagctggaagagctggaagcctcag 180
          |||||||
Db      134622  GCTGGGCTCTTAAGTCTGTGATCCGATGTGAAAAAGCTGGAGGGAAGCTCAG 134563
OY      181  cataccatgtattactttaaaaaacagaaaacagacatgatagtatgtctattttc 240
          |||||||
Db      134562  CATACCATGTATTACTTAAAAACAGAAAAAAGACATGTATGATATCTATTTT 134503
OY      241  ttattattg 249
          |||||||
Db      134502  TTTTATTGG 134494

RESULT  6
AL591464 LOCUS      196374 bp      DNA      linear      HTG 19-DEC-2001
DEFINITION Homo sapiens chromosome 10 clone RP13-302D10, *** SEQUENCING IN
PROGRESS ***; 14 unordered pieces.
ACCESSION  AL591464.2  GI:17973978
VERSION    AL591464
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (sites)
            Burton,J.
            Direct Submission

```

JOURNAL

Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Dec 20, 2001 this sequence version replaced g1:14141513.

Genome Center

Center: Wellcome Trust Sanger Institute
Center code: SC

Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

Project Information

Center project name: DB302D10
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 189117 bases at least Q40
Consensus quality: 192136 bases at least Q30
Consensus quality: 193772 bases at least Q20
Insert size: 221513; sum-of-contigs
Insert size: 221513; 17.0% error; agarose-fp
Quality coverage: 4.57x in Q20 bases; sum-of-contigs quality coverage: 4.33x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 40923: contig of 40923 bp in length
40924 41023: gap of 100 bp
41024 52424: contig of 11401 bp in length
52425 52524: gap of 100 bp
52525 56668: contig of 4144 bp in length
56669 56768: gap of 100 bp
56769 66064: contig of 9296 bp in length
66065 66164: gap of 100 bp
66165 74102: contig of 7938 bp in length
74103 74202: gap of 100 bp
74203 118025: contig of 43823 bp in length
118026 118125: gap of 100 bp
118126 120523: contig of 2398 bp in length
120524 120623: gap of 100 bp
120624 134346: contig of 13723 bp in length
134347 134446: gap of 100 bp
134447 143138: contig of 8692 bp in length
143139 143238: gap of 100 bp
143239 151820: contig of 8582 bp in length
151821 151920: gap of 100 bp
151921 179327: contig of 27407 bp in length
179328 179427: gap of 100 bp
179428 187106: contig of 7679 bp in length
187107 187206: gap of 100 bp
187207 193223: contig of 6017 bp in length
193224 193323: gap of 100 bp
193324 196374: contig of 3051 bp in length.

FEATURES

source

1. 196374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-302D10"
/clone_id="RPC1-13.2"
1. 40923
/note="assembly_fragment:00704
fragment_chain:1"
41024. 52424
/note="assembly_fragment:02076
fragment_chain:1"
52525. 56668
/note="assembly_fragment:02444

fragment_chain:1"
misc_feature 56769..66064
/note="assembly_fragment:00688
fragment_chain:1"
misc_feature 66165..74102
/note="assembly_fragment:01900
fragment_chain:1"
misc_feature 74203..118025
/note="assembly_fragment:02289
fragment_chain:1"
misc_feature 118126..120523
/note="assembly_fragment:01018
fragment_chain:2"
misc_feature 120624..134346
/note="assembly_fragment:01079
fragment_chain:2"
misc_feature 134447..143138
/note="assembly_fragment:00344
fragment_chain:2"
misc_feature 14339..151820
/note="assembly_fragment:01621
fragment_chain:2"
misc_feature 151921..179327
/note="assembly_fragment:01247
fragment_chain:3"
misc_feature 179428..187106
/note="assembly_fragment:01160
fragment_chain:3"
misc_feature 187207..193223
/note="assembly_fragment:01995
fragment_chain:3"
misc_feature 193324..196374
/note="assembly_fragment:02022
clone_end:SP6
vector_side:right"
BASE COUNT 57987 a 41071 c 41893 g 54114 t 1309 others
ORIGIN
Query Match 100.0%; Score 249; DB 2; Length 196374;
Best Local Similarity 100.0%; Pred. No. 2.1e-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cagcgggtgtgcaagaagctcactctgcgtcagatgaagtggtgtggtcgc 60
DB 191512 CAGCGGTGTGCGACGAAGCTACTCTCGCGTCAATGAGTGTGTGTGCTCG 191571
QY 61 gggatcgtgtgtccatcttccttcattgttctgaacatcctgtattgaaccatg 120
DB 191572 GGGATCTGGTGTGCGCCATCTCTTCATTGTTTGGAACATCCTGATGTAAACATG 191631
QY 121 gctgggtgtcctaagtgcctgtgaatcccgatgtggaagaagctggaggtgaagctcag 180
DB 191632 GCTGGGTGTCTTAAAGTGTGGAATCCGATGTGGAAGAACTGGAGCTGAAGCTCAG 191691
QY 181 catcacatgtattcttcttaaaacagaaaaaagacatgtatgtatgtattttt 240
DB 191692 CATCACATGTATTCTTCTTAAACAGAAAAAGACATGTATGTATGTCTATTTT 191751
QY 241 ttattatgg 249
DB 191752 TTTTATTTGG 191760
RESULT 7
AX332939/c AX332939 427 bp DNA linear PAT 09-JAN-2002
LOCUS DEFINITION Sequence 3448 from Patent WO0194629.
ACCESSION AX332939
VERSION AX332939.1 GI:18123573
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens

Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovat, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louie, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Melker, M., Miner, G., Miner, Z., Mitchell, T., Mohabath, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Ogun, M., Okunuga, G., Oraqunye, N., Owiedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tameris, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wood, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 168213)
Moriya, K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:17062822.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GJFC
Center clone name: CH230-98H12

----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to findhaplolist
Consensus quality: 136914 bases at least Q40
Consensus quality: 144233 bases at least Q30
Consensus quality: 150066 bases at least Q20
Estimated insert size: 132352; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-ff estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
5909: contig of 5909 bp in length
5910
6009: gap of unknown length
6010
10545: contig of 4536 bp in length
10546
10645: gap of unknown length
10646
14032: contig of 3387 bp in length
14033
14132: gap of unknown length
14133
17936: contig of 3804 bp in length
17937
18036: gap of unknown length
18037
22114: contig of 4078 bp in length
22115
26249: contig of 4035 bp in length
26250
26349: gap of unknown length
26350
31344: contig of 4995 bp in length

31345
31445
35933: contig of 4489 bp in length
35934
36033: gap of unknown length
36034
40540: contig of 4507 bp in length
40541
40640: gap of unknown length
40641
43879: contig of 3239 bp in length
43880
43980
48194: contig of 4215 bp in length
48195
48294: gap of unknown length
48295
52392: contig of 3998 bp in length
52393
52992: gap of unknown length
52993
54006: contig of 3014 bp in length
54007
55006: gap of unknown length
55007
57239: contig of 1733 bp in length
57240
57339: gap of unknown length
57340
60328: contig of 2889 bp in length
60329
60428: gap of unknown length
60429
63995: contig of 3567 bp in length
63996
64095: gap of unknown length
64096
66845: contig of 2750 bp in length
66846
66945: gap of unknown length
66946
70110: contig of 3165 bp in length
70111
72617: contig of 2407 bp in length
72618
72717: gap of unknown length
72718
75119: contig of 2802 bp in length
75120
75619: gap of unknown length
75620
77931: contig of 2312 bp in length
78031: gap of unknown length
78032
78032: contig of 2794 bp in length
80825
80925: gap of unknown length
80926
83263: contig of 2338 bp in length
83264
83363: gap of unknown length
83364
86126: contig of 2763 bp in length
86127
86226: gap of unknown length
86227
89182: contig of 2956 bp in length
89183
89282: gap of unknown length
89283
90484: contig of 1202 bp in length
90485
90584: gap of unknown length
90585
92580: contig of 1996 bp in length
92581
92680: gap of unknown length
94006: contig of 1326 bp in length
94007
94106: gap of unknown length
94107
96455: contig of 2349 bp in length
96456
96555: gap of unknown length
96556
98594: contig of 2039 bp in length
98595
98694: gap of unknown length
100357: contig of 1663 bp in length
100358
100457: gap of unknown length
100458
102384: contig of 1927 bp in length
102385
102484: gap of unknown length
102485
104202: contig of 1718 bp in length
104203
104302: gap of unknown length
105380: contig of 1078 bp in length
105381
105480: gap of unknown length
105481
107088: contig of 1608 bp in length
107089
107188: gap of unknown length
108957: contig of 1769 bp in length
108958
109057: gap of unknown length
109059
10942: contig of 1885 bp in length
10943
11042: gap of unknown length
11043
113098: contig of 2056 bp in length
113099
113198: gap of unknown length
115082: contig of 1884 bp in length
115083
115182: gap of unknown length
115183
116354: contig of 1172 bp in length
116355
116454: gap of unknown length
116455
118146: contig of 1692 bp in length
118147
118246: gap of unknown length
119751: contig of 1505 bp in length
119752
119851: gap of unknown length
119852
121640: contig of 1789 bp in length
121641
121740: gap of unknown length

Query Match	20.3%;	Score 50.6;	DB 2;	Length 168213;
Best Local Similarity	34.6%;	Pred. No. 0.00091;		
Matches 71; Conservative	0;	Mismatches 134;	Indels 0;	Gaps 0
*	121741	123612:	contig of 1872 bp in length	
*	12613	123712:	gap of unknown length	
*	123713	126065:	contig of 2353 bp in length	
*	126066	126165:	gap of unknown length	
*	126166	127275:	contig of 1110 bp in length	
*	127276	127375:	gap of unknown length	
*	127376	128763:	contig of 1388 bp in length	
*	128764	128863:	gap of unknown length	
*	128864	130267:	contig of 1404 bp in length	
*	130268	130367:	gap of unknown length	
*	130368	131975:	contig of 1608 bp in length	
*	131976	132075:	gap of unknown length	
*	132076	133942:	contig of 1867 bp in length	
*	133943	134042:	gap of unknown length	
*	134043	135320:	contig of 1278 bp in length	
*	135321	135420:	gap of unknown length	
*	135421	136607:	contig of 1187 bp in length	
*	136608	136707:	gap of unknown length	
*	136708	139099:	contig of 2352 bp in length	
*	139100	139199:	gap of unknown length	
*	139200	140405:	contig of 1206 bp in length	
*	140406	140505:	gap of unknown length	
*	140506	142113:	contig of 1608 bp in length	
*	142114	142213:	gap of unknown length	
*	142214	143820:	contig of 1607 bp in length	
*	143821	143920:	gap of unknown length	

Y
43 gltgtgtgtgtggtcgcggatctcgtygtcccatcttccttcaattgtctgaacat 102
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10454 GTGCATGCCCTGTGGCAGACGAGCCCTCCGTTTCCCCCGTTTCATCTGTGGTGTCGAACAT 10513

103 cctgattgttaaacatgctggtgctcaaatgctgtgaatcccgatgtggaataag 162
|||||
10514 CCTGTATTGTAAACCAAGCGCTGGGGGATTAAGNNNNNNNNNNNNNNNNNNNNNN 10573

163 ctgaggtgaaagctcagcataccatgtattactttaaaacagaaaaagacatgta 222
10574 nnn 10533

y 223 tggacatgtctatcttttttttatt 247
 || |||| |
 10634 nnnnnnnnnnnnttccttttgcaat 10658

RESULT 10
C103460

AC103460 168280 bp DNA linear HTG 20-DEC-2
***, 74 unordered pieces. *** SEQUENCING IN PROGRESS

ORGANISM Rattus norvegicus
ERSON AC103460.2 GI:17973302
EYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
NOTES

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 169280)

AUTHORS
Muzly, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Blimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Boute, S., Britton, M., Brown, N., Brown, N., Burnett, N.D., Chabon, C.

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Cullen, M.D., D'Amico, C.D., Davis, C.

Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O.,
Dem, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H.,
Dugan-Rocha, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C.

TITLE	Direct Submission
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 168280)
AUTHORS	Morley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (25-NOV-2001)
	Human Genome Sequencing Center, Department

COMMENT
Dr. Worcester, do remain for college or med school, one
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced g1:17064266.
----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

```
-----
Project information
Center project name: GHOP
Center name: CH230-30D24
-----
Summary Statistics
```

Assembly program:	Phrap; Version 0.990325v11SL	Call
tfndPhrapList		
Consensus quality:	135039 bases at least	Q40
Consensus quality:	141582 bases at least	Q30

Consensus quality:	147182 bases at least Q20
Estimated insert size:	127529; sum-of-contigs estimation
Quality coverage:	0x in Q20 bases; agarose-fp estimation
Quality coverage:	2-3x in Q20 bases; sum-of-contigs estimation

```

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/GenBank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently

```

* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 1000s of Ns, but the exact sizes of the gaps are unknown

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 7204, center of 7204 has its length
* 1

*	7804:	gap of unknown length
*	7705	7705: contig of 5261 bp in length
*	13065	13065: gap of unknown length
*	13066	13066: gap of unknown length
*	13067	13067: gap of unknown length

*	19329	gap of unknown length
*	19330	contig of 5016 bp in length
*	19330	24345
*	24346	24445
*		gap of unknown length


```

repeat_region /rpt_family="Alu"
10008. 10237
/rpt_family="MIR"
repeat_region 10359. 10717
/rpt_family="MALR"
repeat_region 10877. 10962
/rpt_family="MIR"
repeat_region 11116. 11369
/rpt_family="L2"
repeat_region 11370. 11672
/rpt_family="Alu"
repeat_region 11673. 11747
/rpt_family="L2"
repeat_region 11845. 12112
/rpt_family="ERVU"
repeat_region 12113. 12342
/rpt_family="CRL"
repeat_region 12447. 12561
/rpt_family="CRL"
repeat_region 13473. 13601
/rpt_family="MALR"
repeat_region 13629. 13783
/rpt_family="MALR"
repeat_region 16127. 16428
/rpt_family="MERL_type"
repeat_region 16597. 16794
/rpt_family="MIR"
repeat_region 16873. 17103
/rpt_family="MIR"
misc_feature 17766. 18565
/note="similar to EST BG213772 (NID:913735459)"
repeat_region 18021. 18499
/rpt_family="MALR"
repeat_region 18500. 18872
/rpt_family="MALR"
repeat_region 19401. 19561
/rpt_family="MERL_type"
repeat_region 19821. 19877
/rpt_family="CRL"
repeat_region 19881. 20095
/rpt_family="Alu"
repeat_region 20108. 20246
/rpt_family="MERL_type"
repeat_region 20347. 20642
/rpt_family="Alu"
repeat_region 20738. 20817
/rpt_family="Mariner"
repeat_region 21274. 21322
/rpt_family="MIR"
repeat_region 21962. 21985
/rpt_family="(TAA)n"
repeat_region 22291. 22383
/rpt_family="MERL_type"
repeat_region 22725. 22821
/rpt_family="L2"
repeat_region 22837. 23127
/rpt_family="Alu"
repeat_region 24736. 24765
/rpt_family="(CA)n"
repeat_region 25411. 25641
/rpt_family="MIR"
repeat_region 25685. 25922
/rpt_family="Alu"
repeat_region 26039. 26113
/rpt_family="Alu"
repeat_region 27823. 28134
/rpt_family="L2"
repeat_region 28159. 28333
/rpt_family="L2"
repeat_region 29197. 29496
/rpt_family="Alu"
repeat_region 29621. 29668
/rpt_family="AT-rich"

```

```

repeat_region 30581. 30608
/rpt_family="AT-rich"
repeat_region 30805. 30834
/rpt_family="AT-rich"
repeat_region 30967. 31073
/rpt_family="MERL_type"
repeat_region 35404. 35425
/rpt_family="AT-rich"
repeat_region 35508. 35617
/rpt_family="CRL"
repeat_region 36725. 36845
/rpt_family="MIR"
repeat_region 36931. 37008
/rpt_family="MERL_type"

Query Match 15.6%; Score 38.8; DB 9; Length 176741;
Best Local Similarity 62.2%; Pred No. 2.1;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```

```

Oy 150 gatgtgaaagcctgaggtgaaagctcagcatccatgattacttaaaacagaa 209
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98525 GATGAGAGAGGAGGTGAGATGATGCGCTGATGTGTAATATATATATATAT 98466

```

```

Oy 210 aaaaagacatgcatgatactatcttttttttatt 247
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98465 ATATATATATATATATATATATATATATATATATATATATATATATAT 98428

```

RESULT 13

AC023228

LOCUS AC023228 203194 bp DNA linear HTG 01-MAR-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-615C5 map 2, WORKING DRAFT

ACCESSION AC023228

VERSION AC023228.2 GI:7139834

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 203194)

Blumen, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-615C5

Unpublished

2 (bases 1 to 203194)

Blumen, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,

Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,

Choquet, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,

DeBartolano, K., Dewar, K., Domino, M., Doyle, M., Fennestor, J.,

Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,

Gardyna, S., Grant, G., Hagos, B., Heald, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Landers, T., Lehoczy, D., Levine, R., Liu, C., Liu, G., Locke, K.,

Macdonald, P., Marquis, N., McEwan, P., McGuire, A., McKernan, K.,

McPheters, R., Meldrum, J., Menees, L., Morrow, J., Naylor, J.,

Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,

Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rotman, D.,

Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J.,

Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,

Zimmer, A. and Zody, M.

Direct Submission

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 1, 2000 this sequence version replaced gi:6957742.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: http://www-seq.wi.mit.edu


```

Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L6708
Center clone name: 615_C_5
----- Summary Statistics -----
Sequencing Vector: M13: M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 186899 bases at least Q40
Consensus quality: 195828 bases at least Q30
Consensus quality: 198475 bases at least Q20
Insert size: 185000; agarose-fp
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1491: contrig of 1491 bp in length
*
1492 1591: gap of 100 bp
*
1592 2856: contrig of 1265 bp in length
*
2857 2956: gap of 100 bp
*
2957 4454: contrig of 1498 bp in length
*
4455 4554: gap of 100 bp
*
4555 5993: contrig of 1439 bp in length
*
5994 6093: gap of 100 bp
*
6094 8196: contrig of 2103 bp in length
*
8197 8296: gap of 100 bp
*
8297 9731: contrig of 1435 bp in length
*
9732 9831: gap of 100 bp
*
9832 11084: contrig of 1253 bp in length
*
11085 11184: gap of 100 bp
*
11185 12899: contrig of 1715 bp in length
*
12900 12999: gap of 100 bp
*
13000 13527: contrig of 528 bp in length
*
13528 13627: gap of 100 bp
*
13628 15114: contrig of 1487 bp in length
*
15115 15214: gap of 100 bp
*
15215 16378: contrig of 1164 bp in length
*
16379 16478: gap of 100 bp
*
16479 18131: contrig of 1653 bp in length
*
18132 18231: gap of 100 bp
*
18232 20083: contrig of 1852 bp in length
*
20084 20183: gap of 100 bp
*
20184 21117: contrig of 934 bp in length
*
21118 21217: gap of 100 bp
*
21218 24091: contrig of 2874 bp in length
*
24092 24191: gap of 100 bp
*
24192 28023: contrig of 3838 bp in length
*
28030 28129: gap of 100 bp
*
28130 33133: contrig of 5004 bp in length
*
33134 33233: gap of 100 bp
*
33234 37816: contrig of 4583 bp in length
*
37817 37916: gap of 100 bp
*
37917 42224: contrig of 4308 bp in length
*
42225 42324: gap of 100 bp
*
42325 48180: contrig of 5856 bp in length
*
48181 48280: gap of 100 bp
*
48281 54554: contrig of 6274 bp in length
*
54555 54654: gap of 100 bp
*
54655 60966: contrig of 6312 bp in length
*
60967 61066: gap of 100 bp
*
61067 71645: contrig of 10579 bp in length
*
71646 71745: gap of 100 bp
*
71746 81326: contrig of 9581 bp in length
*
81327 81426: gap of 100 bp
*
81427 93995: contrig of 12569 bp in length

```

```

FEATURES
source
*
93996 94095: gap of 100 bp
*
94096 103668: contrig of 9573 bp in length
*
103669 103768: gap of 100 bp
*
103769 113879: contrig of 10111 bp in length
*
113880 113979: gap of 100 bp
*
113980 128120: contrig of 14141 bp in length
*
128121 128220: gap of 100 bp
*
128221 141049: contrig of 12829 bp in length
*
141050 141149: gap of 100 bp
*
141150 155157: contrig of 14008 bp in length
*
155158 155257: gap of 100 bp
*
155258 177981: contrig of 22724 bp in length
*
177982 178081: gap of 100 bp
*
178082 203194: contrig of 25113 bp in length.
Location/Qualifiers
1. 203194
/Organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-615C5"
/clone_lib="RPC1-11 Human Male BAC"
1. 1491
/note="assembly_fragment"
1592. 2856
/note="assembly_fragment"
2957. 4454
/note="assembly_fragment"
4555. 5993
/note="assembly_fragment"
6094. 8196
/note="assembly_fragment"
8297. 9731
/note="assembly_fragment"
9832. 11084
/note="assembly_fragment"
11185. 12899
/note="assembly_fragment"
13000. 13527
/note="assembly_fragment"
vector_end:T7
vector_side:right"
13628. 15114
/note="assembly_fragment"
15215. 16378
/note="assembly_fragment"
16479. 18131
/note="assembly_fragment"
18232. 20083
/note="assembly_fragment"
20184. 21117
/note="assembly_fragment"
vector_end:Sp6
vector_side:right"
21218. 24091
/note="assembly_fragment"
24192. 28029
/note="assembly_fragment"
28130. 33133
/note="assembly_fragment"
33234. 37816
/note="assembly_fragment"
37917. 42224
/note="assembly_fragment"
42325. 48180
/note="assembly_fragment"
48281. 54554
/note="assembly_fragment"
54655. 60966
/note="assembly_fragment"
61067. 71645
/note="assembly_fragment"
71746. 81326
/note="assembly_fragment"

```



```

* 64630 64729: gap of 100 bp
* 64730 69964: contig of 5235 bp in length
* 69965 70064: gap of 100 bp
* 70065 79016: contig of 8952 bp in length
* 79017 79116: gap of 100 bp
* 79117 82147: contig of 3031 bp in length
* 82148 82247: gap of 100 bp
* 82248 92795: contig of 10548 bp in length.

```

FEATURES

SOURCE

```

1. 92795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p34.1-34.3"
/clone="RPI-62B3"
/clone_lib="RPCI-1"
1. 7790
/note="assembly_fragment:00579
fragment_chain:1
clone_end:SP6
vector_side:left"
7891. 19908
/note="assembly_fragment:00853
fragment_chain:1"
20009. 26287
/note="assembly_fragment:00491
fragment_chain:1"
26388. 46199
/note="assembly_fragment:00189
fragment_chain:1"
46300. 54017
/note="assembly_fragment:00445
fragment_chain:1"
54118. 64629
/note="assembly_fragment:00051
fragment_chain:1"
64730. 69964
/note="assembly_fragment:00668
fragment_chain:1"
70065. 79016
/note="assembly_fragment:00709
fragment_chain:1"
79117. 82147
/note="assembly_fragment:00073
fragment_chain:1"
82248. 92795
/note="assembly_fragment:00423
clone_end:T7
vector_side:right"
BASE COUNT 24608 a 19859 c 19994 g 27433 t 901 others
ORIGIN

```

```

Query Match 15.3%; Score 38.2; DB 2; Length 92795;
Best Local Similarity 78.0%; Pred. No. 3;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 189 gtattacttaaaacgaaaaaagacatgtagatgctcatcttttttttatt 247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61779 GTCCTTCTTTAAAAAATAATATATATATATATATATATATATATATATTTT 61837

```

Search completed: September 20, 2002, 05:26:44
 Job time: 5124 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 08:30:59 ; Search time 1856.42 Seconds

(without alignments)
2806.858 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249

Sequence: 1 cagcgcgtgctgcaaggaagc.....gtcatattttttattcgg 249

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	249	100.0	2696	9	AK000761	AK000761 Homo sapi
2	249	100.0	5958	9	AF196185	AF196185 Homo sapi
3	249	100.0	22279	2	U82210	U82210 Homo sapien
4	249	100.0	106198	2	AL160409	AL160409 Human DNA
5	249	100.0	138688	2	AC025823	AC025823 Homo sapi
6	249	100.0	196374	2	AL591464	AL591464 Homo sapi
7	176	70.7	427	6	AX332939	AX332939 Sequence
8	70	28.1	5500	10	AB005549	AB005549 Rattus no
9	50.6	20.3	168213	2	AC103175	AC103175 Rattus no
10	40.8	16.4	168280	2	AC103460	AC103460 Rattus no
11	39	15.7	2305	3	S96842	S96842 SP96=spore-
12	38.8	15.6	176741	9	AC016732	AC016732 Homo sapi
13	38.6	15.5	203194	2	AC023228	AC023228 Homo sapi
14	38.4	15.4	119579	2	AC008425	AC008425 Homo sapi
15	38.2	15.3	92795	2	AL357035	AL357035 Homo sapi
16	38.2	15.3	198965	2	AC103542	AC103542 Rattus no
17	38	15.3	197239	2	AC094815	AC094815 Rattus no
18	37.6	15.1	162260	2	AC021236	AC021236 Homo sapi
19	37.4	15.0	55774	2	AC091303	AC091303 Homo sapi
20	37.2	14.9	76113	2	AC023453	AC023453 Homo sapi
21	37	14.9	125301	3	HS388N15	Z99571 Human DNA s
22	36.8	14.8	12029	3	AE001412	AE001412 Plasmodiu
23	36.8	14.8	185182	9	AC093423	AC093423 Homo sapi
24	36.6	14.7	39608	9	AC009004	AC009004 Homo sapi
25	36.6	14.7	74470	9	AC079404	AC079404 Homo sapi
26	36.6	14.7	79085	9	AC006221	AC006221 Homo sapi
27	36.6	14.7	164309	30	AC067834	AC067834 Homo sapi
28	36.6	14.7	166500	9	AC007785	AC007785 Homo sapi
29	36.6	14.7	169572	2	AC108219	AC108219 Homo sapi
30	36.6	14.7	191830	2	AC025660	AC025660 Homo sapi
31	36.6	14.7	193351	9	AC025656	AC025656 Homo sapi
32	36.6	14.7	205908	2	AL669903	AL669903 Mus muscu
33	36.4	14.6	34722	8	SPBC16C6	AL021767 S.pombe c
34	36.4	14.6	36629	2	AC107279	AC107279 Rattus no
35	36.4	14.6	186135	2	AL669927	AL669927 Mus muscu
36	36.2	14.5	7784	6	AX347326	AX347326 Sequence
37	36.2	14.5	110000	2	AF002753_2	Continuation (3 of
38	36.2	14.5	110000	10	AE008684_1	Continuation (2 of
39	36.2	14.5	131972	4	AC008794	AC008794 Homo sapi
40	36.2	14.5	132090	10	AC005402	AC005402 Mus muscu
41	36.2	14.5	151558	9	AC021766	AC021766 Homo sapi
42	36.2	14.5	155633	2	AC078795	AC078795 Homo sapi
43	36.2	14.5	156060	2	AC004153	AC004153 Plasmodiu
44	36.2	14.5	169546	2	AC004157	AC004157 Plasmodiu
45	36.2	14.5	169794	2	AC004688	AC004688 Plasmodiu

ALIGNMENTS

RESULT	1	
AK000761		
LOCUS	AK000761	2696 bp mRNA linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ20754 fls, clone HEP02246.	
ACCESSION	AK000761	
VERSION	AK000761.1 GI:7021050	
KEYWORDS	oligo capping; fls (full insert sequence).	
SOURCE	Homo sapiens hepatoma cell_line:hepc2 cDNA to mRNA, clone_11b:HEP	
ORGANISM	Clone:HEP02246.	
	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
	1 (sites)	
	Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,	
	Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,	
	Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,	
	Nakamura,Y., Isogai,T. and Sugano,S.	
	NEDO human cDNA sequencing project	
	Unpublished (2000)	
	2 (bases 1 to 2696)	
	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,	
	Shibahara,T., Tanaka,T. and Nakamura,Y.	
	Direct Submission	

JOURNAL

Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology: Shirohane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'-3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by science and Technology Agency).

COMMENT

FEATURES
source

Location/Qualifiers
1..2696
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP02246"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone.lib="HEP"
/note="cloning vector pME18SFL3"
45..1088
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91366.1"
/db_xref="GI:7021051"
/translation="MKAKGMLKGLDMFRFGKHKRDKIEKTKIKIOESTSEER
IMKQDEKTIQAKTEFEREQAREQYAEIIOFHRFGDDELMYGVSSYSMLN
APQSPREGHMADALAYKPKRNSKSPVDSNRSPSNHRIQRIORLOFQOQKQED
VEDRRRTYFEQPMFAPARATOSGRHSVSEVOMOROEERSQOAROVSLPRO
SKNSVSDSDMEQNSPREGPOKSKENPRYSYSGSRNGYIGSGKFNARVYLEPQ
ILROBORREOQMKRPPSEGSNDYSKVDPSIAPPKGPRDVPSPSOVARTLN
RIQEPKRPFTS"

CDS

BASE COUNT 703 a 611 c 695 g 687 t
ORIGIN

Query Match 100.0%; Score 249; DB 9; Length 2696;
Best Local Similarity 100.0%; Pred. No. 1.5e-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 caagcgatgctgagcaagaactcacctcgcgcgtacgtatgagtgctgctgc 60
|||||
2200 CAGCGGTGCTGCGAGGAGACTCTCGCGTACGATTTAGATGTGTGTGGTCTCG 2259
|||||
61 gggatctcggtggtcccactctcctcctcattgtctgaacatcgtlatgtaacatg 120
|||||
2260 GGGATCTCGGTGGCTCCATCTTCTCATTTGTTGGAACATCCTCGATTTGAACCATG 2319
|||||
121 gctggggtgctaaagtgcctgtgaatcccgatgtggaagaagctggaagctcaag 180
|||||
2320 GCTGGGGTCTAAAGTCTGTAATCCGATGTGGAAGAAAGCTGAGGTAAGCTCAG 2379
|||||
181 cataccatgattactttaaaacagaaaaaagacatglatggaatgctatctatctttt 240
|||||
2380 CATACCATGATTACTTTAAAAACAGAAAAAAGACATGATGATGCTATTTT 2439
|||||
241 ttctattgg 249
|||||
2440 TTTTATTGG 2448

RESULT 2

AF196185 5958 bp mRNA linear PRI 16-NOV-2001
LOCUS AF196185
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long
variant mRNA, complete cds.
ACCESSION AF196185
VERSION AF196185.1 GI:13491609
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 5958)
Fang, C.M. and Xu, Y.H.
Down-regulated expression of atypical PKC-binding domain deleted
asp isoforms in human hepatocellular carcinomas
Cell Res. 11 (3), 223-229 (2001)

JOURNAL
AUTHORS

2 (bases 1 to 5958)
Fang, C. and Xu, Y.
Exon/Intron Structure and Splicing Variants of a Novel Human
Polarity Gene, hasip

JOURNAL
AUTHORS

3 (bases 1 to 5958)
Fang, C. and Xu, Y.
Direct Submission
Submitted (18-OCT-1999) Laboratory of Molecular and Cellular
Oncology, Shanghai Institutes for Biological Sciences, Chinese
Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P.R. China

FEATURES

Location/Qualifiers
1..5958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
288..4349
/note="hasip1a"
/product="atypical PKC isotype-specific interacting
protein long variant"
/protein_id="AAK27891.1"
/db_xref="GI:13491610"

CDS

BASE COUNT 1623 a 1395 c 1581 g 1358 t 1 others
ORIGIN

Query Match 100.0%; Score 249; DB 9; Length 5958;
Best Local Similarity 100.0%; Pred. No. 1.5e-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 caagcgatgctgagcaagaactcacctcgcgcgtacgtatgagtgctgctgc 60
|||||
5462 CAGCGGTGCTGCGAGGAGACTCTCGCGTACGATTTAGATGTGTGTGGTCTCG 5521
|||||
61 gggatctcggtggtcccactctcctcctcattgtctgaacatcctglatgtaacatg 120
|||||
5522 GGGATCTCGGTGGCTCCATCTTCTCATTTGTTGGAACATCCTCGATTTGAACCATG 5581
|||||
121 gctggggtgctaaagtgcctgtgaatcccgatgtggaagaagctggaagctcaag 180
|||||

Db	5562	GCATGGCGGTCMAAATGGCTGTGAATCCCGATGTGGAAAAAGCTGAGGTGAACCTCAG	5641
Oy	181	cataccatglatcttacccttaaaacagaaaaaagacatglatggtatglatctttt	240
Db	5642	CATACCACTGATTTACTTTAAAAACAGAAAAAACAATGATGTGATATGCTATTTT	5701
Oy	241	tttatggtg	249
Db	5702	TTTTATTGG	5710
RESULT	3		
LOCUS	U82210	22279 bp	DNA
DEFINITION		Homo sapiens chromosome 10 clone CRI-JC2075 map 10p11.2, ***	linear HTG 30-AUG-2001
ACCESSION	U82210		
VERSION	U82210.1	GI:3983510	
KEYWORDS	HTG; HTGS_PHASe2; HTGS_CANCELED.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.		
TITLE	1 (bases 1 to 22279)		
JOURNAL	Smith,D.R.		
REFERENCE	Sequencing of Human Chromosome 10		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 22279)		
JOURNAL	Du,L. and Smith,D.		
COMMENT	Direct Submission		
	Submitted (13-DEC-1996) Bioinformatics Division, Genome		
	Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA		
	On Dec 9, 1998 this sequence version replaced gi:1773048.		

					* vector sequences.
					* NOTE: This is a 'working draft' sequence. It currently
					* consists of 1 contigs. Gaps between the contigs
					* are represented as runs of N. The order of the pieces
					* is believed to be correct as given, however the sizes
					* of the gaps between them are based on estimates that have
					* provided by the submitter.
					* This sequence will be replaced
					* by the finished sequence as soon as it is available and
				*	the accession number will be preserved.
				1	22279: contig of 22279 bp in length.
FEATURES					
source					Location/Qualifiers
				1..22279	/organism="Homo sapiens"
					/db_xref="taxon:9606"
					/chromosome="10"
					/map="10p11.2"
					/clone="CRI-JC2075"
BASE COUNT	5671 a	5238 c	5397 g	5959 t	14 others
ORIGIN					
	Query Match	100.0%;	Score 249;	DB 2;	Length 22279;
	Best Local Similarity	100.0%;	Pred. No. 1.1e-60;		
	Matches 249; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1	cagcggtggtgcacgaagctcacttcctgcgtcaattatgaagtgtgtgggtccg	60		
Db	10975	CAGCGGTGTTGCACGAAGCTCACTTCGCCGCAATATTAGAAGTGTGGGTCCG	11034		
OY	61	ggagatctcggtgcccatcttccatlcatgtttcttgacaacctgataaccatcy	120		
Db	11035	GGGATCFGGTGCGCTCCCATCTTCCTCATATGTTCGAACAACCCTGTAFTTGAAMC	11094		
OY	121	gcttgtgtgtcctaagaatgctctgtgaatccccgatgttgaaaaagctgtgaagtc	180		
Db	11095	GCTGTGGTGCTAAAGTGCTTGCAATCCCAGATGTGGAAAAAGCTTGAGAGTGAAA	11154		

Accession	Sequence	Position
Oy	101 catcacatgatatcttcctttaaaacagaagaaaaaacatcatgatatgcatctatctttt	240
Db	11155 CATCACATGATATTTACTTTAAAAACAGAAAAACGATGATGATGCTATCTATTTT	11214
Oy	241 ttctatctgg	249
Db	11215 TTTTATTTGG	11223

RESULT	4
AL160409/c	
LOCUS	AL160409 106198 bp DNA linear PRI 18-JUL-2001
DEFINITION	Human DNA sequence from clone RP11-406D17 on chromosome 10, complete sequence.
ACCESSION	AL160409
VERSION	AL160409.12 GI:14970794
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 106198)
AUTHORS	Sycamore,N.
TITLE	Direct Submission
JOURNAL	Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humguery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
COMMENT	On Jul19, 2001 this sequence version replaced gi:14625535.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

```
FEATURES
source
    repeat_region
        /note="MIR repeat: matches 989..1039  
of clone RP11-406D17"
    repeat_region
        /note="MIR repeat: matches 93..144 of consensus"  
1232..1275
    repeat_region
        /note="22 copies 2 mer ac 77% conserved"

```

```
Repeat_region 1344. .1051 /note="AluSx repeat: matches 1. .302 of consensus"
Repeat_region 4229. .4606 /note="MLTJ repeat: matches 1. .403 of consensus"
Repeat_region 4648. .4815 /note="FRAM repeat: matches 4. .171 of consensus"
Repeat_region 5215. .5516 /note="AluSx repeat: matches 1. .300 of consensus"
Repeat_region 6161. .6216 /note="MIR repeat: matches 47. .104 of consensus"
Repeat_region 6217. .6527 /note="AluSx repeat: matches 1. .295 of consensus"
Repeat_region 6528. .6637 /note="MIR repeat: matches 104. .205 of consensus"
Repeat_region 6646. .6950 /note="Aluub repeat: matches 1. .304 of consensus"
Repeat_region 8061. .8354 /note="Aluub repeat: matches 1. .289 of consensus"
Repeat_region 8443. .8602 /note="AluJo/FRAM repeat: matches 152. .308 of consensus"
Repeat_region 10005. .10072 /note="L2 repeat: matches 2673. .2741 of consensus"
Repeat_region 10542. .10849 /note="MER57B repeat: matches 43. .386 of consensus"
Repeat_region 10851. .10975 /note="FLAM.C repeat: matches 15. .133 of consensus"
Repeat_region 11451. .11758 /note="AluJo repeat: matches 1. .312 of consensus"
Repeat_region 12992. .14459 /note="PTR5 repeat: matches 504. .1874 of consensus"
Repeat_region 14460. .14550 /note="MER61E repeat: matches 377. .466 of consensus"
Repeat_region 14551. .14771 /note="PTR5 repeat: matches 1874. .8522 of consensus"
Repeat_region 14772. .14923 /note="AluSx repeat: matches 3. .154 of consensus"
Repeat_region 14930. .15218 /note="Aluub repeat: matches 1. .293 of consensus"
Repeat_region 15223. .15519 /note="AluSg repeat: matches 1. .299 of consensus"
Repeat_region 15661. .15818 /note="MER5A repeat: matches 34. .189 of consensus"
Repeat_region 15868. .15919 /note="26 copies 2 mer aa 73% conserved"
Repeat_region 16133. .16443 /note="AluSx repeat: matches 1. .312 of consensus"
Repeat_region 16599. .16707 /note="L1MC4 repeat: matches 7728. .7834 of consensus"
Repeat_region 16712. .16787 /note="38 copies 2 mer ta 76% conserved"
Repeat_region 16825. .17092 /note="AluJo repeat: matches 1. .292 of consensus"
Repeat_region 17094. .17254 /note="AluJo/FRAM repeat: matches 134. .295 of consensus"
Repeat_region 17579. .17889 /note="AluSg repeat: matches 1. .306 of consensus"
Repeat_region 18807. .19086 /note="MLTJ repeat: matches 3. .238 of consensus"
Repeat_region 19087. .19391 /note="AluSg repeat: matches 1. .310 of consensus"
Repeat_region 19392. .19542 /note="MLTJ repeat: matches 238. .501 of consensus"
Repeat_region 19608. .19904 /note="AluJo repeat: matches 1. .295 of consensus"
Repeat_region 20333. .20642 /note="AluSx repeat: matches 1. .310 of consensus"
Repeat_region 20670. .20956 /note="Aluub repeat: matches 1. .286 of consensus"
Repeat_region 21297. .21601 /note="L2 repeat: matches 2414. .2696 of consensus"
Repeat_region 21857. .22175 /note="AluJo repeat: matches 1. .311 of consensus"
Repeat_region 22502. .22739 /note="AluSx repeat: matches 1. .239 of consensus"
Repeat_region 23135. .23915 /note="L1PA10 repeat: matches 5369. .6158 of consensus"
Repeat_region 27500. .27629 /note="L1PA3 repeat: matches 6171. .6304 of consensus"
Repeat_region 28661. .28984 /note="AluSg repeat: matches 1. .312 of consensus"
Repeat_region 30001. .30132 /note="66 copies 2 mer aa 62% conserved"
Repeat_region 30233. .30328 /note="48 copies 2 mer ta 65% conserved"
Repeat_region 30365. .30644 /note="L1MC4 repeat: matches 6633. .6940 of consensus"
Repeat_region 30645. .31062 /note="MSTB repeat: matches 1. .424 of consensus"
Repeat_region 31063. .31122 /note="L1MC4 repeat: matches 6940. .6995 of consensus"
Repeat_region 31133. .31547 /note="L1MC4 repeat: matches 7167. .7597 of consensus"
Repeat_region 31648. .32167 /note="L1PA6 repeat: matches 5612. .6131 of consensus"
Repeat_region 32275. .32552 /note="AluJo repeat: matches 5. .282 of consensus"
Repeat_region 32592. .32686 /note="L1MC4 repeat: matches 7791. .7972 of consensus"
Repeat_region 33154. .33272 /note="MIR repeat: matches 80. .185 of consensus"
Repeat_region 33283. .33415 /note="AluJo repeat: matches 14. .153 of consensus"
Repeat_region 33416. .34701 /note="PTR5 repeat: matches 172. .2212 of consensus"
Repeat_region 34713. .34840 /note="AluJo/FRAM repeat: matches 178. .302 of consensus"
Repeat_region 35365. .35654 /note="AluJo repeat: matches 1. .288 of consensus"
Repeat_region 36067. .36155 /note="L2 repeat: matches 2622. .2710 of consensus"
Repeat_region 36582. .36804 /note="MER30 repeat: matches 1. .230 of consensus"
Repeat_region 37547. .37852 /note="AluSx repeat: matches 1. .311 of consensus"
Repeat_region 37894. .38050 /note="MER5A repeat: matches 9. .188 of consensus"
Repeat_region 38447. .38744 /note="AluSx repeat: matches 5. .302 of consensus"
Repeat_region 39063. .39193 /note="FLAM.C repeat: matches 1. .131 of consensus"
Repeat_region 39510. .39673 /note="MER5A repeat: matches 4. .189 of consensus"
Repeat_region 40932. .41102 /note="L1PA7 repeat: matches 1497. .1667 of consensus"
Repeat_region 41103. .41726 /note="L1PA7 repeat: matches 5515. .6138 of consensus"
Repeat_region 43386. .43747 /note="MER1B repeat: matches 1. .364 of consensus"
Repeat_region 44459. .44771 /note="AluSg repeat: matches 3. .306 of consensus"
Repeat_region 44991. .45302 /note="AluJo repeat: matches 1. .304 of consensus"
Repeat_region 45536. .51640 /note="L1PA7 repeat: matches 11. .6145 of consensus"
Repeat_region 51663. .51964 /note="Aluub repeat: matches 1. .308 of consensus"
Repeat_region 51966. .52183 /note="109 copies 2 mer aa 69% conserved"
Query Match 100.0%; Score 249; DB 9; Length 106198;
Best Local Similarity 100.0%; Pred. No. 2e-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 cagcggtgtgtgagcgaagctcactcgtcgtcagtttagagtggtgtggttcg 60
DB 88101 CAGCGGTGTGTGAGCAGGAAGCTCACTCTCGCGTCAGTATAGAGTGTGTGTGCTCG 88042
```

Oy	61	ggagactcggtaggcgccacatctctctcaatgcttcctgaacacccgatatgtgaacacatg	120
Db	88041	GGGACTCGTAGGCCTCCCATCTCTCTTCAATGTCTTGAAACATCTGTATTGTGAACCATG	87982
Oy	121	gctgggggtgctaaagtgcctgtgaaatcccgatgctgaaaaagcttgaggctgaagctcag	180
Db	87981	GCTGGGGTGCTAAATGCTGCTGTGATCCGAGTGGAAAAAGCTGGAGGTGAAGCTCAG	87922
Oy	181	cataccatgatatcttaccctttaaacaagaacaaagacatgatatgatatcttattttt	240
Db	87921	CATACCAGTATTTTACTTTAAAAACAGAAAAAAGACATGATGATGATATGCTATTTTTT	87862
Oy	241	tttatatg	249
Db	87861	TTTTTATGG	87853

RESULT	5
LOCUS	AC025823/c
DEFINITION	AC025823 138688 bp DNA linear HTG_24-JAN-2002
ACCESSION	AC025823
VERSION	AC025823.6 GI:13811888
KEYWORDS	HTG; HTGS__PHASE1; HTGS__DRAFT; HTGS__CANCELLED.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 138688) Smith,D.R. Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data Unpublished 2 (bases 1 to 138688) Smith,D.R. Direct Submission Submitted (16-MAR-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA On Apr 27, 2001 this sequence version replaced gi:19887686.
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	1095: contig of 1095 bp in length
1096	1195: gap of unknown length
1196	2203: contig of 1008 bp in length
2204	2303: gap of unknown length
2304	9462: contig of 7159 bp in length
9463	9562: gap of unknown length
9563	25807: contig of 16245 bp in length

FEATURES	source	location/Qualifiers
*	25808	25907: gap of unknown length
*	25908	39871: contig of 13964 bp in length
*	39872	39971: gap of unknown length
*	39972	68325: contig of 28354 bp in length
*	68326	68425: gap of unknown length
*	68425	138688: contig of 70263 bp in length.
FEATURES		
source		
	1..138688	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/chromosome="10"	
	/clone="RP11-60H16"	
	/clone_lib="RPCR-11"	
	1..1095	
misc_feature	/note="assembly_name:Contig14"	
	1196..1203	
misc_feature	/note="assembly_name:Contig15"	
	2304..9452	
misc_feature	/note="assembly_name:Contig55"	
	clone_end:r7"	
	9563..25807	
misc_feature	/note="assembly_name:Contig56"	
	clone_end:SP6"	
	25908..39871	
misc_feature	/note="assembly_name:Contig57"	
	39972..68325	
misc_feature	/note="assembly_name:Contig58"	
	68426..138688	
misc_feature	/note="assembly_name:Contig59"	
	60000 c 29755 g 37221 t	
BASE COUNT	41112 a	30000 c 29755 g 37221 t 600 others
ORIGIN		
Query Match	100.0%	Score 249; DB 2; Length 138688;
Best Local Similarity	100.0%	Pred. No. 2.1e-60;
Matches 249; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
QY	1	cagcgtagtgacgaggaagcctcactcgcgtcagatlaagatggtgtgtgtgtcgcg
	134742	CAGCGTGTGTGGACGAGGACGACCTCGCGCAGCATATTAGAGTGTGTGTGGGCTTCG
QY	61	ggagatcgtgtgtgtcccatcttcctcaatgttcttgacacatcctgtatgttaaccatg
	134682	GGGATCTGTGGTGGCCCATCTTCCTTCATGTTCGTGAACATCCGTATTGTAAACATG
QY	121	gctgggtgtcctaaagtgctgtgaatcccgatgttgaaagaagcttgagatggaagctcag
	134622	GCTGGGTGTCTAAAGTGCCTTGGAATCCGATGTGGAAAGCTGGAGGTGAAGCTCAG
QY	181	catacacgtatttctcttaaaaaacagaaaaaagacatgtatgatatgtctatttct
	134562	CATACACGTATTCTCTTTAAAAAACAGAAAAAACAACATGTATGATATGTATTTTT
QY	241	tttatatgg 249
	134502	TTTTATTGG 134494
RESULT	6	
LOCUS	AL591464	196374 bp DNA linear HTG 19-DEC-2001
DEFINITION	Homo sapiens chromosome 10 clone RP13-302D10, *** SEQUENCING IN	
ACCESSION	AL591464	
VERSION	AL591464.2	GI:17973978
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (sites)	
AUTHORS	Burton,J.	
TITLE	Direct Submission	

NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	1095: contig of 1095 bp in length
1096	1195: gap of unknown length
1196	2203: contig of 1008 bp in length
2204	2303: gap of unknown length
2304	9462: contig of 7159 bp in length
9463	9562: gap of unknown length
9563	25807: contig of 16245 bp in length

	* *	25808	25907:	gap of unknown length
	*	25908	39871:	contig of 13964 bp in length
	*	39872	39971:	gap of unknown length
	*	39972	68325:	contig of 28354 bp in length
	*	68326	68425:	gap of unknown length
	*	68425	138688:	contig of 70263 bp in length.
FEATURES				
SOURCE				
			1..138688	/organism="Homo sapiens"
			/db_xref="taxon:9606"	
			/chromosome="10"	
			/clone="RPJ1-60H16"	
			/clone_lib="RPCR-11"	
			1..1095	
	misc_feature		/note="assembly_name:Contig14"	
			1196..1203	
	misc_feature		/note="assembly_name:Contig15"	
			2304..9462	
	misc_feature		/note="assembly_name:Contig55"	
			clone_end:r7"	
			9563..125807	
	misc_feature		/note="assembly_name:Contig56"	
			clone_end:SP6"	
			25908..139871	
	misc_feature		/note="assembly_name:Contig57"	
			39972..168325	
	misc_feature		/note="assembly_name:Contig58"	
			68426..138688	
	BASE COUNT	41112 a	30000 c	29755 g 37221 t 600 others
ORIGIN				
Query Match			100.0%;	Score 249; DB 2; Length 138688;
Best Local Similarity			100.0%;	Pred. No. 2.1e-60;
Matches	249;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
OY	1	cagcgtagtgacggaaggcaccatctgcgtcaagtataagaigtgtgtggctccg	60	
Db	134742	CAGCGCTGTTGGACGAAGCACCATCTGCCTGTATTAGAAGTGTGTGGGCTTCG	134683	
OY	61	gggatctggtgtgctcccattcttcattatgttttcaaactcatgttaaacatg	120	
Db	134682	GGGATCTGGTGCCGCCCATCTTCCCTCATTTGTGAAACAATCTGTATTGAACCAATG	134623	
OY	121	gctgggtgtcctaagtgctgtgaatccccgattgtgaaaagcttgaaagctcacg	180	
Db	134622	GCTGGGGGTCAAAGTCCCTTGGAATCCCAGATGTGGAAAAACTGGAGGTGAAGCTCAG	134563	
OY	181	catacctgtatttaccttaaacaagaanaaaagacatgttatgatatgtctatttt	240	
Db	134562	CATACCACTGTAATTTCCTTTAAAACAGAAAAAACAACATGTATGATATGTATTTTTT	134503	
OY	241	tttattgagg	249	
Db	134502	TTTTTATTGG	134494	
RESULT	6			
LOCUS	AL591464	196374 bp	DNA	linear HTG 19-DEC-2001
DEFINITION	Homo sapiens chromosome 10 clone RPJ3-302D10, *** SEQUENCING IN			
ACCESSION	AL591464			
VERSION	AL591464.2 GI:17973978			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (sites)			
AUTHORS	Burton,J.			
TITLE	Direct Submission			

[illegible]

Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvach, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtenberg, O., Lien, C., Liu, J., Liu, W., Ma, U., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, J., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okunribido, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and GIBBS, R.

Unpublished
Direct Submission
2 (bases 1 to 168213)
Worley, K.C.

Journal
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:17062822.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GJXC
Center clone name: CH230-98H12

Summary Statistics
Assembly program: Phrap; version 0.990329First call to findPhrapList

Consensus quality: 136914 bases at least Q40
Consensus quality: 144233 bases at least Q30
Consensus quality: 150066 bases at least Q20
Estimated insert size: 132352; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-ff estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 74 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 5909: contig of 5909 bp in length
* 5910 6009: gap of unknown length
* 6010 10545: contig of 4536 bp in length
* 10546 10645: gap of unknown length
* 10646 14032: contig of 3387 bp in length
* 14033 14132: gap of unknown length
* 14133 17936: contig of 3804 bp in length
* 17937 18036: gap of unknown length
* 18037 22114: contig of 4078 bp in length
* 22115 26249: gap of unknown length
* 26249 26349: contig of 4035 bp in length
* 26349 31344: gap of unknown length
* 31344 31344: contig of 4995 bp in length

31345 31444: gap of unknown length
* 31445 35933: contig of 4489 bp in length
* 35934 36033: gap of unknown length
* 36034 40540: contig of 4507 bp in length
* 40541 40640: gap of unknown length
* 40641 43879: contig of 3239 bp in length
* 43880 43979: gap of unknown length
* 43980 48194: contig of 4215 bp in length
* 48195 48294: gap of unknown length
* 48295 52292: contig of 3998 bp in length
* 52293 52392: gap of unknown length
* 52393 55406: contig of 3014 bp in length
* 55407 55506: gap of unknown length
* 55507 57239: contig of 1733 bp in length
* 57240 57338: gap of unknown length
* 57339 60328: contig of 2989 bp in length
* 60329 60429: gap of unknown length
* 60429 63995: contig of 3567 bp in length
* 63996 64095: gap of unknown length
* 64096 66845: contig of 2750 bp in length
* 66846 66945: gap of unknown length
* 66946 70110: contig of 3165 bp in length
* 70111 70210: gap of unknown length
* 70211 72617: contig of 2407 bp in length
* 72618 72717: gap of unknown length
* 72718 75519: contig of 2802 bp in length
* 75520 75619: gap of unknown length
* 75620 77931: contig of 2312 bp in length
* 77932 78031: gap of unknown length
* 78032 80925: contig of 2794 bp in length
* 80926 80925: gap of unknown length
* 80926 83263: contig of 2338 bp in length
* 83264 83363: gap of unknown length
* 83364 86126: contig of 2763 bp in length
* 86127 86226: gap of unknown length
* 86227 89182: contig of 2956 bp in length
* 89183 89282: gap of unknown length
* 89283 90484: contig of 1202 bp in length
* 90485 90584: gap of unknown length
* 90585 92580: contig of 1996 bp in length
* 92581 92680: gap of unknown length
* 92681 94006: contig of 1326 bp in length
* 94007 94106: gap of unknown length
* 94107 96455: contig of 2349 bp in length
* 96456 96555: gap of unknown length
* 96556 98594: contig of 2039 bp in length
* 98595 98694: gap of unknown length
* 98695 100357: contig of 1663 bp in length
* 100358 100457: gap of unknown length
* 100458 102384: gap of unknown length
* 102385 102484: gap of unknown length
* 102485 104202: contig of 1718 bp in length
* 104203 104302: gap of unknown length
* 104303 105380: contig of 1078 bp in length
* 105381 105480: gap of unknown length
* 105481 107088: contig of 1608 bp in length
* 107089 107188: gap of unknown length
* 107189 108957: contig of 1769 bp in length
* 108958 109057: gap of unknown length
* 109059 110942: contig of 1885 bp in length
* 110943 111042: gap of unknown length
* 111043 113098: contig of 2056 bp in length
* 113099 113198: gap of unknown length
* 113199 115082: contig of 1884 bp in length
* 115083 115182: gap of unknown length
* 115183 116354: contig of 1172 bp in length
* 116355 116454: gap of unknown length
* 116455 118146: contig of 1692 bp in length
* 118147 118246: gap of unknown length
* 118247 119751: contig of 1505 bp in length
* 119752 119851: gap of unknown length
* 119852 121640: contig of 1789 bp in length
* 121641 121740: gap of unknown length

24466	30418:	contlig of 5973	bp in length
30419	30518:	gap of unknown length	
30519	35208:	contlig of 4690	bp in length
35209	35308:	gap of unknown length	
35309	39825:	contlig of 4517	bp in length
39826	39925:	gap of unknown length	
39926	42003:	contlig of 5278	bp in length
45204	45303:	gap of unknown length	
45304	46089:	contlig of 2786	bp in length
46090	48189:	gap of unknown length	
48190	53727:	contlig of 5538	bp in length
53728	53827:	gap of unknown length	
53828	57721:	contlig of 3894	bp in length
57722	57821:	gap of unknown length	
57822	60628:	contlig of 2807	bp in length
60629	60728:	gap of unknown length	
60729	65779:	contlig of 2046	bp in length
62775	65874:	gap of unknown length	
62875	65537:	contlig of 2663	bp in length
65538	65637:	gap of unknown length	
65638	67718:	contlig of 1981	bp in length
67619	67718:	gap of unknown length	
67719	71329:	contlig of 3511	bp in length
71230	71329:	gap of unknown length	
71330	73037:	contlig of 1708	bp in length
73038	73137:	gap of unknown length	
73138	76980:	contlig of 3843	bp in length
76981	77080:	gap of unknown length	
77081	77961:	contlig of 2881	bp in length
79962	80061:	gap of unknown length	
80062	82030:	contlig of 1969	bp in length
82031	82131:	gap of unknown length	
82131	84096:	contlig of 1966	bp in length
84097	84197:	gap of unknown length	
84197	85254:	contlig of 1058	bp in length
85255	85354:	gap of unknown length	
85355	87143:	contlig of 1789	bp in length
87144	87243:	gap of unknown length	
87244	89380:	contlig of 2137	bp in length
89381	89481:	gap of unknown length	
89481	91425:	contlig of 1945	bp in length
91426	91525:	gap of unknown length	
91526	93038:	contlig of 1513	bp in length
93039	93138:	gap of unknown length	
93139	95251:	contlig of 2113	bp in length
95252	95351:	gap of unknown length	
95352	96994:	contlig of 1643	bp in length
96995	97094:	gap of unknown length	
97095	99061:	contlig of 1967	bp in length
99062	99161:	gap of unknown length	
99162	101600:	contlig of 2439	bp in length
101601	101700:	gap of unknown length	
101701	104253:	contlig of 2553	bp in length
104254	104353:	gap of unknown length	
104354	106510:	contlig of 2157	bp in length
106511	106610:	gap of unknown length	
106611	106830:	contlig of 1820	bp in length
106831	108530:	gap of unknown length	
108531	110001:	contlig of 1471	bp in length
110002	110101:	gap of unknown length	
110102	111861:	contlig of 1760	bp in length
111862	111961:	gap of unknown length	
111962	113383:	contlig of 1322	bp in length
113384	113383:	gap of unknown length	
113384	116326:	contlig of 2943	bp in length
116327	116426:	gap of unknown length	
116427	117867:	contlig of 1441	bp in length
117868	117967:	gap of unknown length	
117968	119282:	contlig of 1315	bp in length
119283	119382:	gap of unknown length	
119383	120522:	contlig of 1140	bp in length
120523	120622:	gap of unknown length	
120623	121770:	contlig of 1148	bp in length

Query Match	Best Local Similarity	16.4%: Score 40.8; DB 2; Length 168280;
Matches 99; Conservative	50.5%: Pred. NO. 0.57;	Mismatches 97; Indels 0; Gaps 0;
Oy 54	ggtctcgaggatctcggtggtcccatctccatctggtctgaacatcctgtatgta 113	
Db 127130	GGTCATGGTGTGTCATCTGCGGCAATCTTTTCTTTCTTCTTCATATCCCTCGTGTGGAA 127189	
Oy 114	aacctgctgctgggtgcctaaagtgcctgtgatacccgctgtggaaaagctgagtgaa 173	
Db 127190	TGCTATGCAAGAAAGGAAAAACCGCTATTGTGTCAGAGATGATGGAGAAAACCAAGGTCTAA 127249	
Oy 174	agctcagatcacatgattacttacttaaaacagaaaaaagacatgtatgatatgct 233	
Db 127250	CACCTTCTTCTCCCTTTCTCTGATCTACAGATTAAATTAATAAAGGTGGGAGATGTCA 127309	
Oy 234	attttttttattatgg 249	
Db 127310	TGGTGTGTGNGTTGGG 127325	
RESULT 11		
LOCUS	S96842	2305 bp DNA linear INV 07-MAY-1993
DEFINITION	SP96-spore-coat protein [5' region] [Dictyostelium discoideum, prespore cells, Genomic, 2305 nt].	
ACCESSION	S96842	
VERSION	S96842.1	GI:248952
KEYWORDS		
SOURCE	Dictyostelium discoideum prespore cells.	
ORGANISM	Dictyostelium discoideum	
REFERENCE	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.	
AUTHORS	1 (bases 1 to 2305)	
TITLE	Tasaka,M., Hasegawa,M., Nakata,M., Orii,H., Ozaki,T. and Takeuchi,I.	
JOURNAL	Protein binding and DNase-I-hypersensitive sites in the cis-acting regulatory region of the spore-coat SP96 gene of Dictyostelium	
MEDLINE	Mech. Dev. 36 (3), 105-115 (1992)	
REMARK	92239372	
	GenBank staff at the National Library of Medicine created this entry [NCBI gistsq.96842] from the original journal article.	


```

repeat_region /rpt_family="Alu"
10008..10237
repeat_region /rpt_family="MIR"
10359..10717
repeat_region /rpt_family="MaLR"
10877..10962
repeat_region /rpt_family="MIR"
11116..11369
repeat_region /rpt_family="L2"
11370..11672
repeat_region /rpt_family="Alu"
11673..11747
repeat_region /rpt_family="L2"
11845..12112
repeat_region /rpt_family="ERV.L"
12113..12342
repeat_region /rpt_family="CR1"
12447..12561
repeat_region /rpt_family="CR1"
13473..13601
repeat_region /rpt_family="MaLR"
13629..13783
repeat_region /rpt_family="MaLR"
16127..16428
repeat_region /rpt_family="MERL_type"
16597..16794
repeat_region /rpt_family="MIR"
16873..17103
repeat_region /rpt_family="MIR"
17766..18565
misc_feature
18021..18489
repeat_region /rpt_family="MaLR"
18500..18872
repeat_region /rpt_family="MaLR"
19401..19561
repeat_region /rpt_family="MERL_type"
19821..19877
repeat_region /rpt_family="CR1"
19881..20095
repeat_region /rpt_family="Alu"
20108..20246
repeat_region /rpt_family="MERL_type"
20347..20642
repeat_region /rpt_family="Alu"
20738..20817
repeat_region /rpt_family="Mariner"
21274..21322
repeat_region /rpt_family="MIR"
21962..21985
repeat_region /rpt_family="(TAA)n"
22291..22383
repeat_region /rpt_family="MERL_type"
22725..22821
repeat_region /rpt_family="L2"
22837..23127
repeat_region /rpt_family="Alu"
24736..24765
repeat_region /rpt_family="(CA)n"
25411..25641
repeat_region /rpt_family="MIR"
25685..25922
repeat_region /rpt_family="Alu"
26039..26113
repeat_region /rpt_family="Alu"
27823..28134
repeat_region /rpt_family="L2"
28159..28533
repeat_region /rpt_family="L2"
29197..29496
repeat_region /rpt_family="Alu"
29621..29668
repeat_region /rpt_family="AT-rich"

```

```

repeat_region 30581..30608
/rpt_family="AT-rich"
repeat_region 30805..30834
/rpt_family="AT-rich"
repeat_region 30967..31073
/rpt_family="MERL_type"
repeat_region 35404..35425
/rpt_family="AT-rich"
repeat_region 35508..35617
/rpt_family="CR1"
repeat_region 36725..36845
/rpt_family="MIR"
repeat_region 36931..37008
/rpt_family="MERL_type"

Query Match 15.6%; Score 38.8; DB 9; Length 176741;
Best Local Similarity 62.2%; Pred. No. 2.1;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 150 gatgtgaaagctgagtgagtgagctcagcaccatgctattctttaaagcaga 209
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98525 GATGAGAGAGAGCTGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 98466
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 210 aaaaagacatgtagtgatgctattttttttttt 247
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98465 ATATATATATATATATATATATATATATATATATATATATATATATAT 98428
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 13
AC023228
LOCUS AC023228 203194 bp DNA linear HTG 01-MAR-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-615C5 map 2, WORKING DRAFT
SEQUENCE, 32 unordered pieces.
ACCESSION AC023228.2 GI:7139834
VERSION AC023228.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 203194)
Birken, B., Linton, L., Nusbaum, C., and Lander, E.
Homo sapiens chromosome 2, clone RP11-615C5
Unpublished
2 (bases 1 to 203194)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choepe, I., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Feneslor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hages, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, R.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A., and Zody, M.
Direct Submission
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2000 this sequence version replaced gi:6957742.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seg.wi.mit.edu

```

TITLE
 JOURNAL
 COMMENT


```

* 64630 64729: gap of 100 bp
* 64730 69964: contig of 5235 bp in length
* 69965 70064: gap of 100 bp
* 70065 79016: contig of 8952 bp in length
* 79017 79116: gap of 100 bp
* 79117 82147: contig of 3031 bp in length
* 82148 82247: gap of 100 bp
* 82248 92795: contig of 10548 bp in length.

```

FEATURES

```

source
1. .92795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p34.1-34.3"
/clone="RP1-62B3"
/clone_lib="RPCI-1"
1. .7790
/note="assembly_fragment:00579
fragment_chain:1
clone_end:SP6
vector_side:left"
7891. .19908
/note="assembly_fragment:00853
fragment_chain:1"
2009. .26287
/note="assembly_fragment:00491
fragment_chain:1"
26388. .46199
/note="assembly_fragment:00189
fragment_chain:1"
46300. .34017
/note="assembly_fragment:00445
fragment_chain:1"
54118. .64629
/note="assembly_fragment:00051
fragment_chain:1"
64730. .69964
/note="assembly_fragment:00668
fragment_chain:1"
70065. .79016
/note="assembly_fragment:00709
fragment_chain:1"
79117. .82147
/note="assembly_fragment:00073
fragment_chain:1"
82248. .92795
/note="assembly_fragment:00423
clone_end:T7
vector_side:right"
BASE COUNT 24608 a 19859 c 19994 g 27433 t 901 others
ORIGIN

```

```

Query Match 15.38; Score 38.2; DB 2; Length 92795;
Best Local Similarity 78.08; Pred. No. 3;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

OY 189 gtaattactttaaaacagaaaagacatgtagatgctatttttttttatt 247
|| ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 61779 GTCCTTCCTTTTAAAAAATAATATATATATATATATATATTTT 61837

```

Search completed: September 21, 2002, 10:14:35
Job time: 6216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 08:32:09 : Search time 180.75 Seconds
(without alignments)
2365.210 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249
Sequence: 1 cagcgggtgtgagcagaagc.....gtctattttttttattg 249

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	249	100.0	867	22 AAF91875 Human secreted pro
2	249	100.0	5510	22 AAF73337 Human cervical can
3	36.2	14.5	7784	24 ABL34424 Human immune syste
4	36	14.5	2529	18 AAF7468 Hamster Ubiquitin/
5	35.4	14.2	151826	21 AAF22291 BAC containing rep
6	35.2	14.1	619	21 AAF80495 Human colon cancer
7	34.8	14.0	1734	22 AAF17416 Human CDNA sequenc
8	34.8	14.0	2364	21 AAF70246 Plasmodium falcipar
9	34.4	13.8	15416	24 ABL34231 Human immune syste

10	34.4	13.8	15416	24 AAS61453 Human gene regulat
11	34.2	13.7	7784	24 ABL34425 Human immune syste
12	34	13.7	1900	20 AAF25197 Nucleotide sequenc
13	34	13.7	6298	22 AAS45358 Chemically pretrea
14	33.8	13.6	540	22 AAF10129 Human cDNA clone (
15	33.8	13.6	5536	24 ABL32179 Human immune syste
16	33.8	13.6	7784	24 ABL34424 Human immune syste
17	33.6	13.5	13104	19 AAF52167 Streptococcus pneu
18	33.6	13.5	110000	22 AAF84800 Nucleotide sequenc
19	33.4	13.4	1267	22 AAF21012 Bovine-derived DNA
20	33.4	13.4	6306	22 AAK85746 Human immune/haema
21	33.4	13.4	8395	20 AAX13154 Enterococcus faeca
22	33	13.3	881	17 AAT40048 Human tumour necro
23	33	13.3	881	21 AAT28149 Human TR2 receptor
24	33	13.3	1704	19 AAV34509 Human TNF receptor
25	33	13.3	1704	20 AAV83763 Human tumour necro
26	32.8	13.2	265	17 AAT28133 Senescence-related
27	32.8	13.2	8032	24 AAS61408 Human gene regulat
28	32.6	13.1	5914	24 ABL34169 Human immune syste
29	32.6	13.1	6398	23 ABL13526 Drosophila melanog
30	32.4	13.0	5062	22 AAS46966 Cat flea head and
31	32.2	12.9	500	21 AAC95256 DNA encoding novel
32	32.2	12.9	1804	23 AAS83281 S. epidermidis gen
33	32.2	12.9	3069	22 AAF54988 S. epidermidis gen
34	32.2	12.9	3206	22 AAF55018 S. epidermidis gen
35	32.2	12.9	4249	22 AAF54769 Human immune syste
36	32.2	12.9	5536	24 ABL32178 Human immune syste
37	32.2	12.9	6351	24 ABL3745 Human immune syste
38	32.2	12.9	6351	24 ABL34587 Human metastasis a
39	32	12.9	5536	24 ABL32179 Human immune syste
40	32	12.9	17341	21 AAF14872 Genomic DNA sequen
41	32	12.9	32150	22 AAS32454 Human genomic DNA
42	31.8	12.8	407	22 ABA20983 Human nervous syst
43	31.8	12.8	407	22 AAK89467 Human digestive sy
44	31.8	12.8	1383	24 AAS95277 Long terminal repe
45	31.8	12.8	5314	24 ABL32160 Human immune syste

ALIGNMENTS

RESULT	ID	AAFP1875 standard; cDNA: 867 BP.	Human secreted protein-encoding gene 18 cDNA clone HPPFK57, SEQ ID NO:28.
1	AAFP1875		
XX	XX		
AC	AAFP1875;		
XX	XX		
DT	22-MAY-2001 (first entry)		
XX	XX		
DE	Human		
KW	Human; secreted protein; proliferative disorder; cancer; tumour;		
KW	focal abnormality; developmental abnormality; haematopoietic disorder;		
KW	immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;		
KW	inflammation; allergy; neurological disorder; Alzheimer's disease;		
KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;		
KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;		
KW	cardiovascular disorder; angiogenic disorder; kidney disorder;		
KW	gastrointestinal disorder; pregnancy-related disorder;		
KW	endocrine disorder; infection; wound healing; vulnerability;		
KW	cell culture; chemotaxis; food additive;		
KW	binding partner identification; ss.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200118022-A1.		
XX	XX		
PD	15-MAR-2001.		
XX	XX		
PF	31-AUG-2000; 2000WO-US24008.		
XX	XX		
PR	03-SEP-1999; 99US-0152315.		
PR	03-SEP-1999; 99US-0152317.		

Db 5154 ttltattgg 5162

RESULT 3

ABL34424 standard; DNA: 7784 BP.

AC ABL34424;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2397.

XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti anaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antirheumatic; antidiabetic; antiparasitic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN W0200200928-A2.

PD 03-JAN-2002.

PE 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIDEMIOLOGICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

PS Claim 1; SEQ ID NO 2397; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 7784 BP; 1776 A; 134 C; 2219 G; 3655 T; 0 other;

Query Match 14.5%; Score 36.2; DB 24; Length 7784;

Best Local Similarity 77.2%; Pred. No. 1; Mismatches 13; Indels 0; Gaps 0;

Matches 44; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 191 atttacttaaaacagaaacacatgtagatgcatatttttttttttatt 247

Db 5745 atttatt 5801

RESULT 4

AA87468 standard; DNA: 2529 BP.

AC AA87468;

DT 29-JAN-1998 (first entry)

XX Hamster Ubiquitin/S27a 5' untranslated region.

XX CHO; hamster; Ub/S27a; promoter; ubiquitin; primer; ss.

XX Hamster.

XX Key Location/Qualifiers

XX misc_feature 739..2400

XX /tag= a

XX /note= "promoter deletion clone"

XX misc_feature 1545..2400

XX /tag= b

XX /note= "promoter deletion clone"

XX misc_feature 1868..2400

XX /tag= c

XX /note= "promoter deletion clone"

XX misc_feature 2079..2400

XX /tag= d

XX /note= "promoter deletion clone"

XX misc_feature 2195..2400

XX /tag= e

XX /note= "promoter deletion clone"

XX promoter 2130..2243

XX /tag= f

XX misc_feature 813..861

XX /tag= g

XX /note= "homology with amplification sequences"

XX misc_feature 2279..2302

XX /tag= h

XX /note= "polypyrimidine-rich sequence"

XX misc_feature 2374..2392

XX /tag= i

XX /note= "polypyrimidine-rich sequence"

XX misc_signal 2418..2420

XX /tag= j

XX /label= start

XX /note= "start codon"

XX misc_signal 2290

XX /tag= k

XX /note= "transcription start site"

XX misc_signal 2302

XX /tag= l

XX /note= "transcription start site"

XX misc_binding 2238..2243

XX /tag= m

XX /note= "Sp1 binding site"

XX DE19539493-A1.

XX 30-APR-1997.

XX 24-OCT-1995; 95DE-1039493.

XX 24-OCT-1995; 95DE-1039493.

XX 24-OCT-1995; 95DE-1039493.

XX (THOM) THOMAE GMBH KARL.

XX Bergemann K, Enekel B, Gannon F, Noe W;

XX WPI: 1997-246232/23.

XX Nucleic acid containing promoter and regulatory regions of the

XX hamster ubiquitin S27a gene - useful for production of heterologous

XX proteins, especially in CHO cells

XX Claim 4, 25-26; Fig 5; 22pp; German.

XX Cells containing the promoter of the Ub/S27a gene are used for

XX prodn. of heterologous proteins, esp. in Chinese Hamster

XX Ovary cells.

XX The promoter has greater transcriptional activity than the herpes

XX simplex thymidine kinase promoter and about the same as the SV40

CC	promoter.
xx	Sequence 2529 BP; 614 A; 592 C; 641 G; 682 T; 0 other.
sq	

Query Match	14.5%	Score 36	DB 18	Length 2529
Best Local Similarity	50.9%	Pred No. 0.76		
Matches 110	Conservative 0	Mismatches 105	Indels 1	Gaps 1

Oy 29 ggcgacgatataagatcgtctgctcgcccggaattcttcgtgccaccacttccccttc 88
 | ||| ||||| ||||||||| ||| | | ||| ||||| |||
Db 625 gtctgtctgtctgtctgtctgtctgttgt -gtctgtcacggcgcgtctcgatatcatcttac 603

89 atgtgtctgaacatcatctgtattgttaacacatgcttggggtgactaaagtgcctgtgaatcc 148
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 684 ctttgtttaaaaaaatgtttagtccagggttgggtgcactgtgaaagtctgagggttaact 743

Oy 149 cgatgctgnaaaagcgtcgaggygaagctcagcatccagtatttactttaaaaacaga 208
+ + + + + + + + + + + + + + +
Db 744 tgcctgggctcagctcttcctaatagacagaacctcaggtygtcaactcttactgaca 803

```

qy      209 aaaaaagacatgctatgatagtcattttttttc 244
          || | || | || | || | || |
Db      804 gaaccatccaatagccctatctcaatttagtttc 839

```

| RESULT | 5 |
|-------------|------------------------------------|
| AAAF22291/c | |
| ID | AAAF22291 standard: DNA: 151826 BP |

| | | |
|----|-------------|---------------|
| XX | AAF22291; | |
| AC | | |
| XX | | |
| DT | 20-MAR-2001 | (first entry) |

aa BAC containing repeats from centromeres 1-4 #14
De
XX
KW Centromere; michrosome; vector; ds.

Arabidopsis thaliana.
WO200055325-A2.

| | |
|----|-----------------------------|
| PD | 21-SEP-2000. |
| XX | |
| PF | 17-MAR-2000; 2000WO-US07392 |

| | | |
|----|--------------|--------------|
| PR | 18-MAR-1999; | 99US-0125219 |
| PR | 01-APR-1999; | 99US-0127409 |
| PR | 18-MAY-1999; | 99US-0134770 |

PR 17-SEP-1999; 99US-0154603.
XX
PA (UYCH-) UNIV CHICAGO.

PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
v

PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited mitochondria which can serve as vectors for
PT the construction of transgenic plant and animal cells -
PT

PS Claim 102; Page 572-606; 1449pp; English
XX
CC The present invention relates to a recombinant
CC (Arabidopsis thaliana) controllable promoter
CC

CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.

Sequence 151826 BP; 49245 A; 27046 C; 28646 G; 46889 T; 0 other;

| | | | | |
|--------------------------|-------|---------------|----------|---------------|
| Query Match | 14.2% | Score 35.4 | DB 21 | length 151826 |
| Best Local Similarity | 63.5% | Pred. No. 5.3 | | |
| Matches 54, Conservative | 0 | Mismatches 31 | Indels 0 | Gaps 0 |

```
Oy    150 gaatgtgaaaaaacctcgaagtgtaaaagccacgcataccatgatattactttaaaaacaga 209
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   57969 GATTGAAAATAGTPTTTTGTTATAACACATATACGATPTTATNTATTTCCAATCAANA 57910
```

| | | | |
|----|-------|--------------------------|-------|
| QY | 210 | aaanagacatgtatgataatgcta | 234 |
| | | | |
| Db | 57909 | ACAAATTTGTATTAATATTTT | 57885 |

RESULT 6
AAZ80495/c
ID AAZ80495 standard: cDNA: 619 BP

| | |
|----|---------------------------|
| XX | |
| AC | AAZ80495; |
| XX | |
| DT | 07-APR-2000 (first entry) |

| XX | Human colon cancer cell line SW480 cDNA clone SEQ ID NO:579. |
|----|---|
| DE | |
| XX | |
| KW | Human: gene expression product: diagnosis: tumour: colon cancer |

colorectal adenocarcinoma; cell line SW480; cell proliferation;
cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
hyperplasia; ds.

| | |
|----|--------------|
| OS | Homo sapiens |
| XX | |
| PN | W09964576-A2 |
| XX | |

| | |
|----|----------------------------|
| FD | 10 DEC 1999. |
| XX | |
| PF | 09-JUN-1999; 99WO-IB01062. |
| XX | |

XX (FARB) BAYER CORP.
PA
XX

PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE,
PI Schlegel R;
XX

Novel nucleic acids, used to develop products for the diagnosis and treatment of disorders involving unwanted cell proliferation,

XX Claim 15; Page 361; 469pp; English
PS
XX

CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides, which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or altering cancer-related

CC present in a human tumour, particularly in solid tumours, e.g.
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving mutated cell proliferation, such as neoplasia.

| | |
|----|---|
| CC | dysplasia or hyperplasia. |
| XX | |
| SD | Sequence 619 BP; 187 A; 100 C; 103 G; 215 T; 14 other |

| | | | | |
|--------------------------|-------|----------------|-----------|------------|
| Query Match | 14.1% | Score 35.2 | DB 21 | Length 619 |
| Best Local Similarity | 58.1% | Pred. No. 0.79 | | |
| Matches 61; Conservative | 0; | Mismatches 44; | Indels 0; | Gaps 0 |

Oy 139 ctgtgaatcccgatgtygaaaaagctfyaggttgaaagctcagcataccatgttactt 188
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

| | |
|-----------|---|
| CC | Sequence included in the sequence data for this specification and is |
| CC | associated with the human gene regulation-associated genes. |
| CC | Note: The sequence data for this patent did not form part |
| CC | of the printed specification, but was obtained in electronic |
| CC | format directly from WIPO at |
| CC | ftp.wipo.int/pub/published_pcl_sequences |
| XX | |
| SO | Sequence 15416 BP; 4742 A; 239 C; 3068 G; 7367 T; 0 other; |
| | |
| | Query Match 13.8%; Score 34.4; DB 24; Length 15416; |
| | Best Local Similarity 57.4%; Pred. No. 4.5; |
| | Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0 |
| Oy | 140 tgtgatacccgatgttgtaaaagcttgaaggtgaaagctcagacatcattactt 199 |
| | |
| Db | 1966 tttaattttttgtatgtatgaatataataatgatagatagaatggttagtaagt 2025 |
| Oy | 200 aaaaacagaaaaaagacatgtatgatatgtctatttttttttt 247 |
| | |
| Db | 2026 tataaagaataaagtatcatcagaggttttttttttttttttttttttt 2073 |
| RESULT 11 | |
| ID | ABL34425 |
| AC | ABL34425 standard; DNA; 7784 BP. |
| XX | |
| DT | 26-MAR-2002 (first entry) |
| XX | |
| DE | Human immune system associated gene SEQ ID NO: 2398. |
| XX | |
| KW | Human: immune system disease; cytosine methylation; antiasthmatic; |
| KW | antiartherosclerotic; antianemic; cytosolic; nocrotic; |
| KW | neuroprotective; anti-HIV; anticonvulsant; ophthalmological; |
| KW | antirheumatic; antiarthritic; antidiabetic; antipsoriatic; |
| KW | antiflammatory; cancer; eye disease; arteriosclerosis; anaemia; |
| KW | acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; |
| KW | neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; |
| KW | gene; ds. |
| XX | |
| OS | Homo sapiens. |
| PN | WO200200928-A2. |
| XX | |
| PD | 03-JAN-2002. |
| XX | |
| PF | 02-JUL-2001; 2001WO-EP07537. |
| XX | |
| PR | 30-JUN-2000; 2000DE-1032529. |
| PR | 01-SEP-2000; 2000DE-1043826. |
| XX | |
| PA | (EPIG-) EPIGENOMICS AG. |
| XX | |
| PI | Olek A, Piepenbrock C, Berlin K; |
| DR | WPI: 2002-130909/17. |
| XX | |
| PT | Nucleic acid comprising fragment of chemically modified gene, useful |
| PT | for diagnosis and treatment of diseases associated with abnormal |
| PT | cytosine methylation |
| XX | |
| PS | Claim 1; SEQ ID NO 2398; 32pp + Sequence Listing; German. |
| XX | |
| CC | The present invention provides a number of human immune system associated |
| CC | genes which are modified by the methylation of cytosines. The sequences |
| CC | can be used in the diagnosis and treatment of immune system disorders, |
| CC | including eye diseases such as retinopathy, neovascular glaucoma and |
| CC | macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid |
| CC | leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, |
| CC | rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel |
| CC | diseases. The present sequence is a gene of the invention. |

[illegible]

Sequence 6298 BP; 1583 A; 182 C; 1402 G; 3131 T; 0 other;

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAF05003 represent human cDNA sequences and AAU13633 to

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 09:14:55 ; Search time 38.87 Seconds
(Without alignments)
1573.521 Million cell updates/sec

Title: US-09-757-781-21

Sequence: 1 cagcgcgtgtgtgcaggaagc.....gtctattttttttatgtg 249

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/p/odata/2/1na/5A.COMB.seq:*
2: /cgn2_6/p/odata/2/1na/5B.COMB.seq:*
3: /cgn2_6/p/odata/2/1na/6A.COMB.seq:*
4: /cgn2_6/p/odata/2/1na/6B.COMB.seq:*
5: /cgn2_6/p/odata/2/1na/PC/US.COMB.seq:*
6: /cgn2_6/p/odata/2/1na/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------|--------------------|
| 1 | 36 | 14.5 | 2529 | US-09-051-969A-5 | Sequence 5, Appli |
| 2 | 35.2 | 14.1 | 619 | US-09-328-111-579 | Sequence 579, App |
| 3 | 31.4 | 12.6 | 579 | US-09-146-950-3 | Sequence 3, Appli |
| 4 | 31.4 | 12.6 | 591 | US-09-146-950-19 | Sequence 19, Appli |
| 5 | 31.4 | 12.6 | 1596 | US-09-146-950-17 | Sequence 17, Appli |
| 6 | 31.4 | 12.6 | 1724 | US-08-509-024-1 | Sequence 1, Appli |
| 7 | 31.4 | 12.6 | 1724 | US-09-333-279-1 | Sequence 1, Appli |
| 8 | 31.4 | 12.6 | 1929 | US-09-146-950-1 | Sequence 1, Appli |
| 9 | 31.4 | 12.6 | 4622 | US-08-509-024-6 | Sequence 6, Appli |
| 10 | 31.4 | 12.6 | 4622 | US-09-333-279-6 | Sequence 6, Appli |
| 11 | 30.4 | 12.2 | 2646 | US-08-539-304A-5 | Sequence 5, Appli |
| 12 | 30.4 | 12.2 | 2663 | US-08-136-743B-3 | Sequence 3, Appli |
| 13 | 30.4 | 12.2 | 3252 | US-09-118-442-1 | Sequence 1, Appli |
| 14 | 30.4 | 12.2 | 3252 | US-09-677-064-1 | Sequence 1, Appli |
| 15 | 29.8 | 12.0 | 1724 | US-07-847-010-18 | Sequence 18, Appli |
| 16 | 29.6 | 11.9 | 860 | US-08-773-608A-1 | Sequence 1, Appli |
| 17 | 29.6 | 11.9 | 2658 | PCT-US95-08295-22 | Sequence 22, Appli |
| 18 | 29.4 | 11.8 | 524 | US-08-460-739-1 | Sequence 1, Appli |
| 19 | 29.4 | 11.8 | 4875 | US-08-460-739-1 | Sequence 1, Appli |
| 20 | 29.2 | 11.7 | 19124 | US-08-487-826B-13 | Sequence 13, Appli |
| 21 | 29 | 11.6 | 279 | US-08-686-878A-50 | Sequence 50, Appli |
| 22 | 29 | 11.6 | 279 | US-08-721-489-4 | Sequence 4, Appli |
| 23 | 28.8 | 11.6 | 684 | US-08-577-463A-1 | Sequence 3, Appli |
| 24 | 28.8 | 11.6 | 2946 | US-09-175-928-3 | Sequence 3, Appli |
| 25 | 28.4 | 11.4 | 2622 | US-08-766-014-23 | Sequence 23, Appli |
| 26 | 28.4 | 11.4 | 2902 | US-08-714-918-95 | Sequence 95, Appli |
| 27 | 28.4 | 11.4 | 2902 | US-09-265-315-95 | Sequence 95, Appli |

| | | | | | | |
|------|------|------|------|---|-------------------|--------------------|
| c 28 | 28.4 | 11.4 | 2902 | 4 | US-09-265-315-95 | Sequence 95, Appli |
| c 29 | 28.4 | 11.4 | 2902 | 4 | US-09-265-417-95 | Sequence 95, Appli |
| c 30 | 28.4 | 11.4 | 5300 | 1 | US-08-766-014-1 | Sequence 1, Appli |
| c 31 | 28.2 | 11.3 | 618 | 4 | US-08-961-810-17 | Sequence 17, Appli |
| c 32 | 28.2 | 11.3 | 618 | 4 | US-08-352-903D-17 | Sequence 17, Appli |
| c 33 | 28.2 | 11.3 | 2075 | 1 | US-08-238-163-3 | Sequence 3, Appli |
| c 34 | 28 | 11.2 | 392 | 1 | US-08-469-421-8 | Sequence 8, Appli |
| c 35 | 28 | 11.2 | 392 | 1 | US-08-250-975-8 | Sequence 8, Appli |
| c 36 | 28 | 11.2 | 392 | 2 | US-08-605-002A-8 | Sequence 8, Appli |
| c 37 | 28 | 11.2 | 392 | 2 | US-08-950-449A-8 | Sequence 8, Appli |
| c 38 | 28 | 11.2 | 392 | 5 | PCT-US94-10529-8 | Sequence 8, Appli |
| c 39 | 28 | 11.2 | 2612 | 4 | US-09-105-390-7 | Sequence 7, Appli |
| c 40 | 28 | 11.2 | 6124 | 4 | US-08-213-419B-3 | Sequence 3, Appli |
| c 41 | 27.8 | 11.2 | 886 | 1 | US-08-469-427A-1 | Sequence 1, Appli |
| c 42 | 27.8 | 11.2 | 886 | 2 | US-08-609-443B-1 | Sequence 1, Appli |
| c 43 | 27.8 | 11.2 | 886 | 2 | US-08-569-063C-1 | Sequence 1, Appli |
| c 44 | 27.8 | 11.2 | 1002 | 4 | US-08-960-780-43 | Sequence 43, Appli |
| c 45 | 27.8 | 11.2 | 1002 | 4 | US-09-073-898-43 | Sequence 43, Appli |

ALIGNMENTS

RESULT 1
US-09-051-969A-5
Sequence 5, Application US/09051969A
Patent No. 6063598
GENERAL INFORMATION:
APPLICANT: ENEKEL, BARBARA
APPLICANT: GANNON, FRANK
APPLICANT: BERGMANN, KLAUS
APPLICANT: NOE, WOLFGANG
TITLE OF INVENTION: INTENSIVE HOMOLOGOUS PROMOTER OBTAINED
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,969A
FILING DATE: 1998-09-30
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FLESHNER, RAZ E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 0652.1690000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2529 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2418..2465
US-09-051-969A-5
Query Match 14.5%; Score 36; DB 3; Length 2529;
Best Local Similarity 50.9%; Pred. No. 0.042;

| | | | | |
|-----------------------|----------------|--|----------|------------|
| Query Match | 14.1% | Score 35.2 | DB 4 | Length 619 |
| Best Local Similarity | 58.1% | Pred. No. 0.039 | | |
| Matches 61 | Conservative 0 | Mismatches 44 | Indels 0 | Gaps 0 |
| QY | 139 | ctgtgaatcccgatgtggaanaagcttgagtgatgaaagctcaagcatcacatgatatctactt | 198 | |
| | | | | |
| Db | 472 | cagtaaaactctgggttttagatgcattctgtagaagaagtaaggaactgtcagttatcttttc | 413 | |
| QY | 199 | taaaaaacgaaaaaaaagacatgatagtatgatatgtcattttttttt | 243 | |
| | | | | |
| Db | 412 | tttaaacancagaaaaaacgatatattggcagatgatgtgtttt | 368 | |

RESULT 3
 US-09-146-950-3/c
 ; Sequence 3, Application US/0914650A
 Patent No. 6287808

| | Query Match | Best Local Similarity | 12.6%
51.0% | Score 31.4 | DB 4 | Length 591 |
|----|-------------|--|----------------|------------|---------|------------|
| | Matches | 74; Conservative | 0; Mismatches | 71; Indels | 0; Gaps | 0 |
| QY | 53 | gggctcggggatctcgttggtcccaatcttcattgttctgaacatcctgattgt | 112 | | | |
| Db | 481 | gggtgtctctgactctcggtgctccctcttctgcaccccttgagcccggtgaggtggcgt | 422 | | | |
| QY | 113 | aaacacatgagtcggtgtgttaagtgcctgtgaatccgcatgtggaanaaagtcggaggtga | 172 | | | |
| Db | 421 | AAGGGGGGACGCGGCGGAGTGGTCCCGCTCTGGACGATGAGAAATGGCTGGGGCTGC | 362 | | | |
| QY | 173 | aagctcagcatatcatgatttact | 197 | | | |
| Db | 361 | AGCCACACACGGCGCTTCTGTCTCT | 337 | | | |

RESULT 5
US-09-146-950-17/c
; Sequence 17, Application US/09146950A

```

; Patent NO. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-146-950-17

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1596;
Best Local Similarity 51.0%; Pred. No. 1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

QY 53 gggctcggggatcctcggtggtcccatcttccttcattgttctgaacatccgtatgt 112
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 587 GGGTCTCGACTCTGCGTCCCTCTTCTGCAACCTCTGGCCGGGCTGGAGTGGCGT 528
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 113 aaacatggtcgtggtgctaaagtgcctgtgaatcccgatgtgaaaaagctgaggtga 172
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 527 AAGCGGGGACAGCGGCGAGTGGTCCCGCTGACAGATGAGAAAGTGGCTGGGCTGC 468
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 aagtcagcatatcattatct 197
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 467 AGCCACACACGCGCTTCTGTCT 443
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
US-08-509-024-1/c
; Sequence 1, Application US/08509024B
; Patent No. 6291207
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/08/509,024B
; CURRENT FILING DATE: 1995-07-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-509-024-1

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1724;
Best Local Similarity 51.0%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

QY 53 gggctcggggatcctcggtggtcccatcttccttcattgttctgaacatccgtatgt 112
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 774 GGGTCTCGACTCTGCGTCCCTCTTCTGCAACCTCTGGCCGGGCTGGAGTGGCGT 715
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 113 aaacatggtcgtggtgctaaagtgcctgtgaatcccgatgtgaaaaagctgaggtga 172
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 AAGCGGGGACAGCGGCGAGTGGTCCCGCTTCTGACAGATGAGAAAGTGGCTGGGCTGC 655
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 aagtcagcatatcattatct 197
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 AGCCACACACGCGCTTCTGTCT 630
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
US-09-333-279-1/c

```

```

; Sequence 1, Application US/09333279
; Patent No. 6303336
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/09/333,279
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-333-279-1

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1724;
Best Local Similarity 51.0%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

QY 53 gggctcggggatcctcggtggtcccatcttccttcattgttctgaacatccgtatgt 112
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 774 GGGTCTCGACTCTGCGTCCCTCTTCTGCAACCTCTGGCCGGGCTGGAGTGGCGT 715
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 113 aaacatggtcgtggtgctaaagtgcctgtgaatcccgatgtgaaaaagctgaggtga 172
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 AAGCGGGGACAGCGGCGAGTGGTCCCGCTTCTGACAGATGAGAAAGTGGCTGGGCTGC 655
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 aagtcagcatatcattatct 197
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 AGCCACACACGCGCTTCTGTCT 630
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
US-09-146-950-1/c
; Sequence 1, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (297)...(875)
; US-09-146-950-1

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1929;
Best Local Similarity 51.0%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

QY 53 gggctcggggatcctcggtggtcccatcttccttcattgttctgaacatccgtatgt 112
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 777 GGGTCTCGACTCTGCGTCCCTCTTCTGCAACCTCTGGCCGGGCTGGAGTGGCGT 718
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 113 aaacatggtcgtggtgctaaagtgcctgtgaatcccgatgtgaaaaagctgaggtga 172
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 717 AAGCGGGGACAGCGGCGAGTGGTCCCGCTTCTGACAGATGAGAAAGTGGCTGGGCTGC 658
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 aagtcagcatatcattatct 197
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 657 AGCCACACACGCGCTTCTGTCT 633
    ||||| ||| ||||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The University of Pennsylvania
; STREET: Suite 330
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19104-3246
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,743B
; FILING DATE: 10/14/93
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 3957-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5459063e
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2663 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-136-743B-3

Query Match      12.2%; Score 30.4; DB 1; Length 2663;
Best Local Similarity 63.9%; Pred. No. 2.8;
Matches 46; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 176 ctcaagatccatgattcttcttaaaacagaaaaagacatgctatgatatctat 235
Db 2301 cttatataaaatgacatcactatataatataatataatataatataatataat 2242
Qy 236 ttttttttact 247
Db 2241 attttttttttt 2230

RESULT 13
US-09-118-442-1/C
; Sequence 1, Application US/09118442B
; Patent No. 6197561
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; APPLICANT: Wang, Xun
; APPLICANT: Bowen, Benjamin A.
; TITLE OF INVENTION: Genes Controlling Phylate Metabolism in
; FILE REFERENCE: 0706
; CURRENT APPLICATION NUMBER: US/09/118,442B
; EARLIER FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/055,446
; EARLIER FILING DATE: 1997-08-11
; EARLIER APPLICATION NUMBER: 60/055,526
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: 60/053,944
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1

```

```

; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)...(2666)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3252)
; OTHER INFORMATION: n = A,T,C or G
; US-09-118-442-1

Query Match      12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 92 gtctgacatccctgattgtaaaccaagctgggtgctaaagtgctgtaaccgca 151
Db 1862 gttcaaggcactcttaatatagacgattcttgagcgacacactcctgttagaagaagct 1803
Qy 152 tgtgaaaagctggaagtgaaagctcagcatcacatgtlactttaaaacagaaa 211
Db 1802 tggctgaatgctgaacgaattgctcatcaagcttcttaagctcactgaaaaactctga 1743
Qy 212 aaagacatgctatgatatgtctattttttt 243
Db 1742 taatagctgcctcaatttttcaattttctttt 1711

RESULT 14
US-09-677-064-1/C
; Sequence 1, Application US/09677064
; Patent No. 6291224
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; TITLE OF INVENTION: Genes Controlling Phylate Metabolism in
; FILE REFERENCE: 0706D
; CURRENT APPLICATION NUMBER: US/09/677,064
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)...(2666)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3252)
; OTHER INFORMATION: n = A,T,C or G
; US-09-677-064-1

Query Match      12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 92 gtctgacatccctgattgtaaaccaagctgggtgctaaagtgctgtaaccgca 151
Db 1862 gttcaaggcactcttaatatagacgattcttgagcgacacactcctgttagaagaagct 1803
Qy 152 tgtgaaaagctggaagtgaaagctcagcatcacatgtlactttaaaacagaaa 211
Db 1802 tggctgaatgctgaacgaattgctcatcaagcttcttaagctcactgaaaaactctga 1743
Qy 212 aaagacatgctatgatatgtctattttttt 243
Db 1742 taatagctgcctcaatttttcaattttctttt 1711

```


QY 152 tctggaagaagctggaggtgaaagctcagcatatgatttactttaaacagaaaa 211
DB 1802 TGGTGGCAATAGGTGAAGAAATGTCATCAAACTTTGTAGCTCAGTGAATAACTCTCGA 1743
QY 212 aaagacatgtatgatatgtctatctttttt 243
DB 1742 TAATAGCTGCCTCAATTTTCAATTTCTCTTT 1711

RESULT 15

PCT-US96-12374-1/c
; Sequence 1, Application PC/RUS9612374
; GENERAL INFORMATION:
; APPLICANT: Northwestern University
; TITLE OF INVENTION: Herpes Virus Entry Mediator
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Milnamow & Katz, Ltd.
; STREET: 180 N. Stetson, Suite 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/12374
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Northrup, Thomas E.
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: NOR3446P020PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5400
; TELEFAX: (312) 616-5460
; TELEX: --
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1724 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 294..1145
; NAME/KEY: mat_peptide
; LOCATION: 294..1142
; PCT-US96-12374-1

Query Match 12.0%; Score 29.8; DB 5; Length 1724;

Best Local Similarity 50.3%; Pred. No. 3.5;

Matches 73; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 53 gggctcgggagatcgtggtccatcttcctcatgttctgaacatccgtatgt 112
DB 774 GGGTGTCTGCTGACTGCTGCTCCTCTGCAACCTTGCGCCGGGCTGAGGTGCGCT 715
QY 113 aaacatgtcgtgggtgctaaagtctgtaatcccgatgtgaaaaaagctgaggtga 172
DB 714 AACGGCGGACAGCGGCGCAGTGGTCCCGTCTGAGAGATGACAGAGTGGCTGGGCTGC 655
QY 173 aagctcagcatcacatgtatttact 197
DB 654 AGCCACACACAGCGGTCTCTGTCT 630

Search completed: September 21, 2002, 10:17:28
Job time: 3753 sec

536 bp mRNA linear EST 29-SEP-1998
 xl Soares fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
 IMAGE:1674370 3', mRNA sequence.
 .1 GI:3415789

 iens
 Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 s 1 to 536)
 http://www.ncbi.nlm.nih.gov/ncicgap.
 Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Index
 ned (1997)
 Robert Strausberg, Ph.D.
 gapbs-r@mail.nih.gov
 is available royalty-free through LNL; contact the
 smorium (info@mag.hnl.gov) for further information.
 length: 1447 Std Error: 0.00
 er: -40mJ3 fwd. ET from Amer sham
 ility sequence stop: 445.
 Location/Qualifiers
 1. 536
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1674370"
 /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
 /sex="male"
 /dev_stage="20 week post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site-1: Pac I; Site-2: Eco RI;
 This is a subtracted version of the original Soares fetal
 liver spleen 1NFLS library. 1st strand cDNA was primed
 with a Pac I - 0190(OT) primer 15'
 AACCTGCAAGATTTAAATTAAGAATCTTTTTTTTTTTTTTTT 3'),
 double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

174 a 126 c 114 g 122 t

Query Match 100.0%; Score 249; DB 9; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.1e-51;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagcggatggtgaggaagcctcactcgcgtcagatagagtggtggtctcg 60
|||||
DB 497 CACGGGTGGTGGCAGAACTCACTCGGCTAGTATTAGAGTGTGTGGTCTCG 438
|||||
OY 61 ggagctcgtggtcctccatctcctcattgttcgaaacccgtattgtaaccatg 120
|||||
DB 437 GGGATCTCGGTGGCTCCCATCTTCCTCATTTGTGAACATCCTGTATTCTAAACATG 378
|||||
OY 121 gctggggtgctaaagtgcctgtaatccgagtgtggaagaagctggaggtgaaagctcag 180
|||||
DB 377 GCGGGGTGCTAAAGTGCCTGTAAATCCGATGTGAAAAAGCTGGAGTGAAGCTCAG 318
|||||
OY 181 catcacatgtattactttaaaacagaaaaagacatgtatgatatgtctattttt 240
|||||
DB 317 CATACCATGATTACTTTTAAACAGAAAAAAGACATGATGATGATGCTATTATTTT 258
|||||
OY 241 ttattatgg 249
|||||
DB 257 TTTTATTGG 249

RESULT 2

LOCUS

AI497808/c

DEFINITION

tm89f02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2165307 3'

ACCESSION

AI497808

VERSION

AI497808.1

KEYWORDS

EST.

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
This clone is available royalty-free through LNL; contact the
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bbrp/image/image.html
Insert Length: 779 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 468.
Location/Qualifiers
1. 536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2165307"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"

FEATURES

source

/note="Organ: brain; Vector: p773D-Pac (Pharmacia) with a
modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTACCAATCTGAAGTGGAGCGCCGCAATGAGGTGTTTTTTTTTTTTTTT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT

174 a 126 c 114 g 122 t

Query Match 100.0%; Score 249; DB 9; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.1e-51;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagcggatggtgaggaagcctcactcgcgtcagatagagtggtggtctcg 60
|||||
DB 497 CACGGGTGGTGGCAGAACTCACTCGGCTAGTATTAGAGTGTGTGGTCTCG 438
|||||
OY 61 ggagctcgtggtcctccatctcctcattgttcgaaacccgtattgtaaccatg 120
|||||
DB 437 GGGATCTCGGTGGCTCCCATCTTCCTCATTTGTGAACATCCTGTATTGTAACATG 378
|||||
OY 121 gctggggtgctaaagtgcctgtaatccgagtgtggaagaagctggaggtgaaagctcag 180
|||||
DB 377 GCTGGGTGCTAAAGTGCCTGTGAATCCGATGTGAAAAAGCTGGAGTGAAGCTCAG 318
|||||
OY 181 catcacatgtattactttaaaacagaaaaagacatgtatgatatgtctattttt 240
|||||
DB 317 CATACCATGATTACTTTTAAACAGAAAAAAGACATGATGATGATGCTATTATTTT 258
|||||
OY 241 ttattatgg 249
|||||
DB 257 TTTTATTGG 249

RESULT 3

LOCUS

AI038061/c

DEFINITION

ox21b05.x1 Soares fetal_liver_spleen_INFIS_S1 Homo sapiens cDNA

clone IMAGE:1656945 3', mRNA sequence.

ACCESSION

AI038061

VERSION

AI038061.1

KEYWORDS

EST.

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1466 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 465.
Location/Qualifiers
1. 561
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1656945"
/clone_lib="Soares_fetal_liver_spleen_INFIS_S1"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;

FEATURES

source

FEATURES
source

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-RC3-H10585-160>)
300-022-d046t3=2000-03-16&t4=1)
Seq primer: puc 18 forward


```

/obj_xref="taxon:9606"
/clone_image="1931593"
/clone_idb="Soares_NhNMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below): Vector: pRTT3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbM, pregnant uterus
NBH9U, and fetal heart NBH19W) were mixed, and ss cicles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cdnas from pools
of 5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      163 a      112 c      100 g      111 t
ORIGIN

```

| | | | | |
|---------------------------|---------|--------------------|-------|-------------------|
| Query Match | 95.2%; | Score 237; | DB 9; | Length 486; |
| Best Local Similarity | 100.0%; | Pred. NO. 1.1e-48; | | |
| Matches 237; Conservative | 0; | Mismatches | 0; | Indels 0; Gaps 0; |

| | | | |
|----|-----|---|-----|
| OY | 13 | caggaaagctcaatcccgacgtaattgaagcgttcttgtagttccggagatcccgtgtg | 72 |
| Db | 486 | CAGGAAGCTCACTTCGCGCTCAGTATTAAAGTGtGTGGGTCCGGGAAATCCGGTG | 422 |
| OY | 73 | gtcccaacttccttcatacttgtttcttgaacctcctgtattgtaaacatatgctgggtcta | 13 |
| Db | 426 | GCTCCCACTTCCTTCATATTGTCTGAACACTCTGTATTSTAAACCATGGCTGGGGTCCA | 36 |
| OY | 133 | aagtgcctcgtgatccgcatctgtgaaaaaacctctgagsgtgaagaatcgaatcacatgttat | 19 |
| Db | 366 | AAGTGCCTCGTGAATCCCGATGTGGAAAAAAGCTGGAGGTTAAAGCTCACAGCATACCATGTAAT | 30 |
| OY | 193 | ttaactttaaaaacagaaaaaaaagaacatctatgatatgtctatctttttttttatgtg | 249 |
| Db | 306 | TTACTTTAAAACAGAAAAAAGAACAATGATGCAATGTCTATTTTTTTTTTAAATGG | 250 |

| | | | | | |
|------------|----------------|---------------|--------------|--------------------------|-----------------|
| RESULT | 8 | | | | |
| LOCUS | A1952086/c | | | | |
| DEFINITION | A1952086 | 486 bp | mRNA | linear | EST 06-SEP-1999 |
| | wx46g11.x1 | NCI_CGAP_Lu28 | Homo sapiens | cDNA clone IMAGE:2546756 | 3' |
| | mRNA sequence. | | | | |

| | | |
|----------|--------------|------------|
| VERSION | AT952086.1 | GI:5744396 |
| KEYWORDS | EST. | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--------------------|--|--|--------------------|-----------------------------------|
| 1 (bases 1 to 486) | NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap . | National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index | Unpublished (1997) | Contact: Robert Strausberg, Ph.D. |

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: strausberg@u.washington.edu

Information can be found through the I.M.A.G.E. Consortium/ILNUL at: www.bio.llnl.gov/dbrrp/image/image.html
Seq primer: -40UP from Glbco
High quality sequence stop: 413.

FEATURES
source

| BASE COUNT | ORIGIN |
|------------|-------------------|
| 164 a | 112 c 100 g 110 t |

```

Query Match: 95.2% Score 237 DB 9 Length 486
Best Local Similarity 100.0% Pred. No. 1,1e-48;
Matches 237, Conservative 0, Mismatches 0, Indels 0, Gaps 0

```

| | | | |
|----|-----|--|------|
| QY | 13 | caaaagctcactctgcgcgaagaagaatgctgtgagctcgtcgtggatcctggtg | 72 |
| Db | 486 | CAGNAGCTCACCTCGCGTCAGATTAGATGCTGTGTGGGCTCGGGGATTCGGTG | 4227 |
| QY | 73 | gtcccaatcttccttcatalgtctcgaacatcctglatgtaaacatcgcgcgtggtccta | 1323 |
| Db | 426 | GCTCCATCTTCCTTCAATFTTTCGAACTCCCTTATTGTAAACATGCGCTGGGCTCTA | 367 |
| QY | 133 | aagtgctctggaatcccgatgttgaaaaagcttggaagtgaagctcagcataccatglat | 192 |
| Db | 366 | AAGTGCTGTGAATCCCGATGTGGAAAAAGCTGGAAGTGGAAGCTCAGCATACATGTAT | 307 |
| QY | 193 | ttaacttaaaaaagaaaaaagcagatgtagtatagtcattttttttattatgg | 249 |
| Db | 306 | TTACTTTAAAAACGAAAAAAGCATGTATAGGATATCTATTTTTTTTTTATTTGG | 250 |

| | |
|------------|---|
| RESULT | 9 |
| A112988/c | |
| LOCUS | 545 bp mRNA linear EST 27-OCT-1998 |
| DEFINITION | gc5c0-c3.x1 Soares_pregnant uterus_NBPu Homo sapiens cDNA clone IMAGE:171502 3 , mRNA sequence. |
| ACCESSION | A112988 |
| VERSION | A112988.1 GI:3598502 |

| ORGANISM | REFERENCE |
|---|--------------------|
| <i>Homo sapiens</i> | 1 (pages 1 to 545) |
| Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | |

| | |
|---------|---|
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP) |
| JOURNAL | Tumor Gene Index |
| COMMENT | Unpublished (1997) |
| | Contact: Robert Strausberg, Ph.D. |

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE consortium (info@image.llnl.gov) for further information.
Insert Length: 1140 Std Error: 0.00
Seq primer: -40m3 fwd. Et from Amersham
High quality sequence stop: 444.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .545 |

```

/clone_11b="Soares_pregnant_uterus_NbHPv"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AAGTGAAGAATTCGCGCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

```



```

/adb_xref="taxon:9606"
/clone_image="2016909"
/clone_id="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: p773D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with Not I - oligo(dT) primer [5'
TGTTACCACTGTGAAGTGGGAGCGCGCATAGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

```

| Query Match | Similarity | 86.3% | Score 215 | DB 9 | Length 463 |
|-------------|------------|--|-------------------|------------|------------|
| Best Local | Similarity | 100.0% | Pred. No. 3.1e-43 | | |
| Matches | 215 | Conservative | 0 | Mismatches | 0 |
| | | | | Indels | 0 |
| | | | | Gaps | 0 |
| OY | 35 | glaataagatgltgltgltgltgltcgcgggagatcgtgtgctccatctctctcatgtt | 94 | | |
| Db | 463 | GTATTAGAGTGTGTGTGTGGGTCTGGGGGATCTCGGTGGCTCCATCTTCTTCAATTGT | 404 | | |
| OY | 95 | ctgaacatccgttatcttgaacatagcgtggtgtcctaagtgcctgtgaaaccgatgt | 154 | | |
| Db | 403 | CTGAACATCTGTATTGTGAACCATAGCGCTGGGGTCTAAAGTCCGTGTGAATCCCGANGT | 344 | | |
| OY | 155 | ggaaaaagctgtgaagtgaaagctcagcatatcatgttacttctaaaaacgaaaaaa | 214 | | |
| Db | 343 | GGAAAAGCTGGAGGTGAAAGCTGCAGCATACATGTATTATTAAAAACGAAAAAAA | 284 | | |
| OY | 215 | gacatgatgatgatgtctattttttttatttgc | 249 | | |
| Db | 283 | GACATGATGATGATGATCTATTATTTTTTTTATTTGG | 249 | | |

```

/clone="IMAGE:1913978"
/clone_1lb="NCI_CGAP_Ki45"
/tissue_type="2_pooled_tumors (clear cell type)"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACGGAGAAGATTGCGCGGCCCATATTTTCTTTTCTTTT 3'],
(double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "

```

| | | | | |
|-----------------------|--------|---|--------|--------------|
| Query Match | 83.5% | Score 208 | DB % | Length 462 |
| Best Local Similarity | 100.0% | Pred. No. 1,7e+41 | | |
| Matches | 208 | Conservative | 0 | Mismatches 0 |
| | | | Indels | 0 |
| | | | Gaps | 0 |
| Oy | 42 | agctgctgctgtggtgcttcgggagacctcggtgagctccactcttcctcatgtgtcgaaca | 101 | |
| Db | 462 | AGTGTGTGTGTGGGCTCTCGGGGATTCCTCGGTGCTCCCATCTTCATTTGTTGGAACA | 403 | |
| Oy | 102 | tcctctatctgaaacacatgagctgggggtgctaaagtgcctgtgaatcccgatctggaanaa | 161 | |
| Db | 402 | TCTCTTATTGTGAACCATAGGCTGGGGTGTCTAAAGTGTCTGTGAATCCGATGTGAAAAA | 343 | |
| Oy | 162 | gctggaagtgaaagctcagcataccatgtatttactttaaanaacagaaaaaagacatgtc | 221 | |
| Db | 342 | GCTGAGGCTGGAAGCTCAGCATACCATGTATTACTTTAAAAACAGAAAAAAGACATCT | 283 | |
| Oy | 222 | atggaatgctcatatttttttttttattg | 249 | |
| Db | 282 | ATGCAATATGCTCATTTTTTTTTTTATATGG | 255 | |

| FEATURES | source |
|------------|---|
| LOCUS | AI030934/c |
| DEFINITION | AI030934 : 462 bp mRNA linear EST 01-FEB-1999 |
| ACCESSION | g01e02.x1 NCI-CGAP_kid5 Homo sapiens cDNA clone IMAGE:1913978 3' |
| VERSION | AI030934 |
| KEYWORDS | mRNA sequence. |
| SOURCE | AI030934.1 GI:4004805 |
| ORGANISM | human. |
| REFERENCE | Homo sapiens |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| TITLE | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index |
| JOURNAL | Unpublished (1997) |
| COMMENT | Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1094 Std Error: 0.00
Seq primer: -400p from Glbco
High quality sequence stop: 425.
location/Qualifiers
1. 462
/organism="Homo sapiens"
/db_xref="taxon:9606" |

```

RESULT 13
AL369277/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

```



```

/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: ovary; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCATCTGAAGTGGAGCGCGCGGTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Felima Bonaldo."

```

```

BASE COUNT      147 a      89 c      88 g      103 t
ORIGIN

```

```

Query Match      70.7%: Score 176; DB 9; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.5e-33;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      74  ctccatcttccttcattgtctgaacacccgtatgttaaccatgctgaggtgctaa 133
      |||||||
Db      427  CTCCCATCTTCTCTCATTTGTTCTGAACATCCTCTATTGTAAACCATGCGTGGGCTGCTAA 368
      |||||||
Oy     134  agtgccctgtaatcccgatgtgaaagctggaggtgaagctcagcatatgtat 193
      |||||||
Db     367  ACTGCGCTGTGAATCCGATGTGAAAGCTGAGGTGAAGCTCAGCATACCATGTATT 308
      |||||||
Oy     194  tactttaaaacagaaaaaagacatgtatgtatgtctattttttttattgg 249
      |||||||
Db     307  TACTTTAAAAACGAAAAAAGACATGATATGCTATTTTCTTTTATTGG 252
      |||||||

```

```

Search completed: September 21, 2002, 09:42:10
Job time: 5937 sec

```